IMPROVING TRAFFIC SIGNAL PERFORMANCE USING HIGH-RESOLUTION DATA

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

HENG HU

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

ADVISOR: HENRY X. LIU

DECEMBER 2013

Acknowledgements

I would like to express my deepest appreciation to my adviser, Dr. Henry X. Liu, without whom this dissertation would not have been possible. His continuous support and guidance on this research have helped me overcome so many difficulties. He is not only a good advisor in research and study, but also a great mentor in life.

Many thanks go to Dr. Gary A. Davis. I have learned a lot from him, especially the statistical techniques in transportation engineering. I also greatly appreciate the help from Dr. John Hourdos. His constructive suggestion about simulation and modeling made this research more solid. Special thanks go to Dr. Shuzhong Zhang for his service in the committee. His insightful comments in mathematical formulation greatly improve this dissertation. Finally, I would like to thank Dr. David M. Levinson, who taught me how to think engineering problems from an economical point of view.

This research could not have been done without the tremendous support from the Minnesota Department of Transportation. I would like to offer my special thanks to Mr. Steve Misgen, Mr. Timothy Bangsund, Mr. Curt Krohn and Mr. Ron Christopherson.

I would also like to thank my fellow colleagues and friends, Dr. Xinkai Wu, Dr. Xiaozheng He, Dr. Guizhen Yu, Mr. Jie Sun and Mr. Jianfeng Zheng for their help in this research and in my life.

Last but not least, my heartfelt appreciation goes to my family, especially my wife, Grace Huang, Thank you all for the continuous support and for sharing the pains and joys in the past several years. I am so proud of you.

To my families

Abstract

Traffic congestion has become a more and more severe problem for metropolitan areas all over the world. Although lots of effort has been devoted to improve the operation on major corridors, how to efficiently and effectively manage traffic based on the existing infrastructure is still a challenging task, and the task seems even more challenging for signalized arterials due to lack of traffic monitoring and data collection system. This research aims to improve the traffic signal performance based on the collected high-resolution traffic signal data and the derived performance measures by the SMART-Signal system developed at the University of Minnesota. In particular, this research focuses on the following three areas:

1) **Optimize offsets to reduce congestion**

Traditionally, offset optimization for coordinated traffic signals fails to consider the stochastic nature of field traffic. Using the archived high-resolution traffic signal data, in this research, we develop an arterial offset optimization model which will take two well-known problems with vehicle-actuated signal coordination into consideration: the early return to green problem and the uncertain intersection queue length problem. To account for the early return to green problem, we introduce the concept of conditional distribution of the green start times for the coordinated phase. To handle the uncertainty of intersection queue length, we adopt a scenario-based approach that generates optimal offsets using a series of traffic demand scenarios as the input to the optimization model. Both the conditional distributions of the green start times and traffic demand scenarios can be obtained from the archived high-resolution traffic signal data. Under different traffic conditions, queues formed by side-street and main-street traffic are explicitly considered in the derivation of intersection delay. The objective of this model is to minimize total delay for the main coordinated direction and at the same time it considers the performance of the opposite direction. The results from the field implementation

show that the proposed model can reduce travel delay of coordinated direction significantly without compromising the performance of the opposite approach.

2) Manage oversaturated signal arterials

Under oversaturated traffic conditions, signal timings need to be adjusted accordingly in order to alleviate the detrimental impacts caused by oversaturation. In this research, our focus is to mitigate two types of detrimental effects, signal phase failure with residual queue and downstream queue spillover. Building upon the previous work on the oversaturation severity indices, a maximum-flow based approach to manage oversaturated intersections is developed. The proposed model maximizes the discharging capacity along oversaturated routes, while satisfying the constraints on available green times. We show that a simple forward-backward procedure (FBP) can be used to obtain the optimal solution to the maximum flow model. The forward process aims to increase green time to mitigate oversaturation, therefore improve the throughput for the oversaturated approach; and the backward process aims to gate the traffic at some intersections to prevent residual queues and downstream queue spill-back when the available green time is insufficient. The algorithm is tested using a microscopic traffic simulation model for an arterial network in the City of Pasadena, CA. The results indicate the model can effectively and efficiently reduce oversaturation and improve system performance.

3) Mange integrated corridors

An integrated control model is proposed to manage traffic congestion along a freeway and a parallel signalized arterial. This model focuses on freeway diversion, which aims to utilize available capacities along parallel arterial routes to reduce network congestion. The potential impact of the diverting traffic to the performance of the arterial route is considered in this research and the maximum flow based signal control model is utilized to manage congestion on the arterial route. The integrated control model does not need the time-dependent traffic demand information as most of previous approaches do and it

is suitable for online applications because of its low computation burden. The model is tested using the microscopic traffic simulation in the I-394 and TH 55 corridor in Minneapolis, MN. The results indicate that the model can effectively and efficiently reduce network congestion.

Table of Contents

List of Tables ..... **ix**

List of Figures ..... **x**

Notations ..... **xii**

1. Introduction **1**

1.1 Problem Statement 1

1.2 Research Contributions 2

1.3 Dissertation Organization 4

2. The SMART Signal System **5**

2.1 Background 5

2.2 Data Collection Units 5

2.3 Performance Measures 7

3. Arterial Offset Optimization Using Archived High-resolution Traffic Signal Data **8**

3.1 Background 8

3.2 High-resolution Traffic Signal Data 11

3.2.1 Conditional Distribution of Green Start Times 12

3.2.2 Traffic Scenarios 16

3.3 Model Formulation 19

3.3.1 Scenario-based Offset Optimization 19

3.3.2 Delay Calculation 22

3.4 Solution Method and Optimization Results 32

3.5 Field Evaluation 35

3.6 Summary 38

4. Managing Oversaturated Signal Arterials: A Maximum Flow Based Approach 40

4.1 Background 40

4.2 OSI-based Mitigation Strategies for Single Intersection 42

4.3 Maximum Flow Problem Formulation for Arterial Oversaturation 47

4.3.1 Control Variables 48

4.3.2 Constraint Analysis 50

4.3.3 Maximum Flow Based Control Model 53

4.3.4 Model Discussion 58

4.4 Solution Method – A Forward-Backward Procedure (FBP) 60

4.4.1 FBP for One Oversaturated Route 61

4.4.2 FBP for a Network with Two Intersecting Oversaturated Routes 64

4.4.3 Optimality Analysis of the FBP 65

4.5 Simulation Test 69

4.5.1 Simulation Design 69

4.5.2 Test Results 71

4.5.3 Sensitivity Analysis 75

4.6 Summary 76

5. An Integrated Control Model for Managing Network Congestion 77

5.1 Background 77

5.2 Problem statement 79

5.3 Model formulation 79

5.3.1 Performance estimation 80

5.3.2 Diversion Control 84
5.4 Case study and simulation 87
5.5 Summary 93
6. Conclusion and Future Research 95
6.1 Conclusion 95
6.2 Future Research 96
References 98

List of Tables

Table 3-1 Parameters for GA 33
Table 3-2 Eastbound and Westbound Input volume (7:00AM~9:00AM) comparison between 9/3/2009 and 9/14/2009 36
Table 3-3 Side street Input volume (7:00AM~9:00AM) comparison between 9/3/2009 and 9/14/2009 36
Table 3-4 Eastbound and Westbound average delay comparison between 9/3/2009 and 9/14/2009 38
Table 4-1. Traffic flow conditions during the simulation period 70
Table 4-2. Control strategy comparison 71
Table 4-3. Network performance comparison 72
Table 4-4. Comparisons of two-hour throughputs 73
Table 4-5. Network performance comparison between strategies with estimation errors 76
Table 5-1. Network performance comparison 92
Table 5-2. Network performance comparison with demand variations 93

List of Figures

Figure 2-1 SMART-Signal data collection units 6
Figure 2-2 SMART-Signal data transmission..... 6
Figure 2-3 SMART-Signal data tables 7
Figure 3-4. Installation of the SMART-SIGNAL system on TH55, Golden Valley, MN (Source: Google Maps) 11
Figure 3-5 Distribution of green start time of Phase 2 in local clock, TH 55 & Boone Ave, 7:00 AM ~ 9:00 AM, 6/15/2009 ~ 6/26/2009 10 weekdays 13
Figure 3-6 Dual-ring control diagram of intersection TH 55 & Boone Ave during weekday AM peaks 13
Figure 3-7 Dual Ring Control diagram of an example intersection 14
Figure 3-8 Phase 2 (Eastbound through) green start time vs. associated number of actuations $\rho_{1,2}$, TH 55 & Boone Ave, 6/15/2009 ~ 6/26/2009 10 weekdays 16
Figure 3-9 Conditional distributions of Phase 2 (EB) green start time $P(G_{1,2} \mid \rho_{1,2})$,TH 55 & Boone Ave, 6/15/2009 ~ 6/26/2009 10 weekdays 16
Figure 3-10 Boundary input volumes and turning percentages 18
Figure 3-11 Empirical cumulative distributions for eastbound and westbound input volume 18
Figure 3-12 Data processing flow chart 22
Figure 3-13 Queuing process between two intersections when $d_n^c \geq 0, d_n^g \geq 0$ 27
Figure 3-14 Queuing process between two intersections when $d_n^c < 0, d_n^g \geq 0$ 29
Figure 3-15 Queuing process between two intersections considering arrival profiles within a cycle when $d_n^c \geq 0, d_n^g \geq 0$ 31
Figure 3-16 GA results for each generation 34
Figure 3-17 Coordination result for both field and optimized offsets 35
Figure 3-18 Estimated EB queue length based on different offset settings 38

Figure 4-1 Green extension for Scenario 1 44
Figure 4-2 Red extension for Scenario 2 45
Figure 4-3 Red reduction (at downstream intersection) for Scenario 3 46
Figure 4-4 Intersections along an oversaturated arterial 47
Figure 4-5 Red Time Changes & Green Time Changes 49
Figure 4-6 Signal timing changes ($\Delta r_{n,i}(t) < 0,\ \Delta g_{n,i}(t) > 0$) 50
Figure 4-7 An example of applying strategies 2 & 3 to eliminate spillover 51
Figure 4-8 Maximum flow network for one oversaturated route 55
Figure 4-9 Two intersecting oversaturated routes 56
Figure 4-10 Maximum flow network for two intersecting oversaturated routes 58
Figure 4-11 VISSIM simulation network (a) normal flow condition (b) increased flow condition 70
Figure 4-12 Comparison of network performance 72
Figure 4-13 Southbound throughputs over time under different strategies 73
Figure 4-14 Comparison of side streets' maximum queue length in each cycle 74
Figure 5-1 Problem statement 79
Figure 5-2 Diversion control from freeway to arterial 85
Figure 5-3 Case study site: the TH 55/I-394 corridor, Minneapolis, MN (Source: Google Maps) 88
Figure 5-4 Flow-density diagram from three detectors at segment 4 88
Figure 5-5 VISSIM network of the TH 55/I-394 corridor 89
Figure 5-6 Demand profiles for the simulation period 90
Figure 5-7 Travel times of general route and diverting route under different scenarios 91
Figure 5-8 Travel time and diversion rate 92

Notations

n	Index of intersection
i	Index of phase
$\rho_{n,i}$	Associated number of vehicle actuations, for coordinated phase i of intersection n
$V_{n,i}$	Cycle input volume for phase i of intersection n
$G_{n,i}$	Green start time for phase i of intersection n
$P\left(G_{n,i} \mid \rho_{n,i}\right)$	Conditional distribution of green start time given associated number of actuations $\rho_{n,i}$
$U_{n,d}$	Cycle input volume of direction d at intersection n
$\varphi_{n,d}^{x}$	Turning percentages of direction d at intersection n
S	Set of traffic scenarios
D	Total travel delay of coordinated directions
O	Vector of intersection offsets
Ω	Feasible set of O
α	Weighting term representing the importance of the minor coordinated direction
β	Trimming factor
J_{β}	Number of scenarios left after trimming
r_n	Red light duration at intersection n
W_n^s	Cycle input volumes from side streets to downstream at intersection n
W_n^m	Cycle input volumes from main street to downstream at intersection n

q_n^s	Arrival flow rate per lane to downstream through links formed by side street traffic of intersection *n*
q_n^m	Arrival flow rate per lane to downstream through links formed by main street traffic of intersection *n*
$q_n(\tau_n)$	Arrival flow rate per lane to downstream through links at any second τ_n of a cycle
$w_{n,d}^x(\tau_n)$	Discharging flow rate of the corresponding movements of intersection *n* at any second τ_n
z_n	Number of lanes of the coordinated direction at intersection *n*
d_n^c	Time difference of cycle starts between intersection n and downstream intersection $n+1$
d_n^g	Time difference of green starts between intersection n and downstream intersection $n+1$
t_n^q	The portion of green at intersection *n* within which discharged vehicles will cause a queue at downstream
q_n^d	Discharging rate of intersection n
TOSI	Temporal Oversaturation Severity Index, caused by the residual queue that creates the detrimental effect in temporal dimension
SOSI	Spatial Oversaturation Severity Index, caused by the spillover that creates the detrimental effect in spatial dimension
$TOSI_{n,i}$	*TOSI* value at intersection *n* for phase *i*
$SOSI_{n,i}$	*SOSI* value at intersection *n* for phase *i*
$g_{n,i}$	Green time duration at intersection *n* for phase *i*.
$\Delta g_{n,i}$	Green time adjustment at intersection *n* for phase *i*
$\Delta r_{n,i}$	Red time adjustment at intersection *n* for phase *i*
$S_{n,i}(t)$	Unusable green time caused by spillover (i.e., *SOSI* > 0) at intersection *n* for phase *i* during time period of *t*
$T_{n,i}(t)$	unusable green time caused by residual queue (i.e., *TOSI* > 0) at intersection *n* for phase *i* during time period of *t*
$g_{n,i}^a(t)$	Available green time for intersection *n* and phase *i* during time period of *t*
$Z_{n,i}$	Set of conflicting phases to phase *i* at intersection *n*

$Q_{n,p}^{\max}(t)$	Maximum queue size per lane for phase *p* at intersection *n* at control interval t
$s_{n,p}$	Saturation flow rate per lane for phase *p* of intersection *n*
$f_{n,i}^{r}(t)$	Input flow for red reduction in the maximum flow network
$f_{n,i}^{g}(t)$	Input flow for green extension in the maximum flow network
$\varepsilon_{n,i}^{S}(t)$	Slack variable representing the unsolved portion of $S_{n,i}(t-1)$ in control period *t*
$\varepsilon_{n,i}^{T}(t)$	Slack variable representing the unsolved portion of $T_{n+1,i}(t-1)$ in control period *t*
$\Delta r_{n,i}^{F}(t)$	Red time adjustment at intersection *n* for phase *i* in the forward process
$\Delta g_{n,i}^{F}(t)$	Green time adjustment at intersection *n* for phase *i* in the forward process
$R_{n,i}(t)$	Residual capacity (in seconds) at intersection *n* for phase *i*
$\Delta g_{i}^{B}(t)$	Green time adjustment for phase *i* in the backward process
$\Delta_{I,i}^{R}(t)$, $\Delta_{I,j}^{R}(t)$	Requested green time increase for phase *i* and *j* of intersection *I*
$k_m(t)$	Average density of segment *m* during control period *t*
$\theta_m(t)$	Average occupancy of segment *m* during the control period *t*
$v_m(t)$	Average speed of segment *m* during the control period *t*
L_v	Average vehicle length
L_d	Length of loop detector
$V_m^{30}(t)$	Average 30-second volume of segment *m* during control period *t*
$T^f(t)$	Travel time along the freeway corridor during control period *t*
$T^a(t)$	Travel time along the signalized arterial corridor during control period *t*
$T^{f\to a}(t)$	Diversion cost from freeway to arterial during control period *t*

l_m^f	Length for segment *m* of the freeway corridor
$Q_{n,i}(\tau)$	Queue length for phase *i* at intersection *n* at given second τ
$l_{n,i}^a$	Link length of phase *i* at intersection *n*
$v_{n,i}^a$	Free flow speed of phase *i* at intersection *n*
$\eta_{n,i}(t)$	Residual capacity (in vehicles) of intersection n for the phase *i* during the control period t
$\gamma_{n,i}(t)$	Average cycle discharging volume for phase *i* of intersection n during control period t
$\phi(t)$	Diversion rate from freeway to arterial at control period *t*
$\pi^{f\to a}(t)$	Diverted traffic volume from freeway to arterial in control period *t*
$\psi_{n,i}(t)$	Residual queue caused by the increase of traffic (due to diversion) for phase *i* of intersection n in control period *t*
$\Delta\lambda_{n,i}(t)$	Increase of arrival traffic due to diversion control for phase *i* of intersection n in control period *t*

1. Introduction

1.1 Problem Statement

Due to the increasing traffic demand but limited facility capacity, traffic congestion has become a more and more severe problem for metropolitan areas not only in the United States but also around the world. How to efficiently and effectively manage traffic along major corridors appears to be an urgent task for researchers and practitioners. Particularly, the problem becomes even more challenging on signalized arterials due to the lack of traffic monitoring and data collection system. As indicated by the 2012 National Traffic Signal Report Card (http://www.ite.org/reportcard/), the "Management and operation of traffic signals" was graded as 69. It shows a little progress comparing with a grade of 65 in 2007, however, a great deal of effort is strongly needed to improve the operation on signalized arterials.

The development on the SMART-Signal (**S**ystematic **M**onitoring of **A**rterial **R**oad **T**raffic and **Signal**) system makes it possible to continuously monitor the traffic situation of signalized arterials in a quantitative way. In particular, the system can simultaneously collect event-based high resolution traffic data from signalized intersections and generate performance measures in real time. The collected high-resolution traffic signal data and the derived performance measures have opened a new window to improve traffic signal operation. Based on those, this research aims to optimize arterial signal control parameters and develop new control strategies to reduce traffic congestion. Two major themes are the focus of this dissertation:

1) Offset optimization to reduce congestion

Offset settings are critical to the operation of coordinated signals. Traditionally, offset optimization for coordinated traffic signals is based on average travel times between intersections and average traffic volumes at each intersection, without consideration of the stochastic nature of field traffic. However, there are two problems associated

with vehicle-actuated coordinated traffic signals: (1) the early return to green problem because of the traffic flow variations on non-coordinated approaches, (2) the variation of waiting queue lengths at coordinated directions. However, none of the previous studies have addressed the two problems identified above simultaneously. The two problems are intertwined together, as traffic demand variations will lead to uncertain green start times, and uncertain green start times will result in stochastic length of waiting queues even with deterministic demands. Thus, the two problems need to be considered simultaneously.

2) Real-time control to manage oversaturated arterials

It is well-known that managing oversaturated arterials requires significantly different approaches for signal timing from those used for under-saturated conditions. Over the past several decades, plenty of literature has been devoted to this research area. However, in reality, none of these approaches have been widely accepted or implemented because of the following reasons. Firstly, most of the proposed models require time-dependent traffic demand information, which cannot be provided by the existing detection system. Secondly, it is almost impossible, especially for real-time traffic control, to solve large optimization programs with hundreds, or even thousands of decision variables and constraints. Third, many practitioners do not feel comfortable to implement the "black-box" method provided by most of the optimization models. Therefore, a simple but effective control model to deal with oversaturation on signalized arterials needs to be developed.

1.2 Research Contributions

This thesis makes contributions in the areas of arterial signal optimization, arterial oversaturation management and integrated corridor control. Specifically:

1) **It proposes a data-driven approach to optimize offsets for vehicle-actuated coordinated traffic signals**, using the massive amount of signal status and

vehicle actuation data collected from the field. The proposed approach, for the first time, takes the two problems of vehicle-actuated traffic signal coordination into consideration simultaneously. To account for the early return to green problem, for the coordinated phase, we introduce the concept of conditional probability of the green start times. To handle the uncertainty of intersection queue length, we adopt a scenario-based optimization approach. The proposed model is expected to reduce travel delay of coordinated direction without compromising the performance of the opposite approach.

2) **It develops a simple but effective control model to deal with oversaturation on signalized arterials.** The control model is not based on time-dependent demand information, as most of the previous studies do. It is built upon the characterization of oversaturation discussed by Wu et al. (2010), in which, oversaturation is quantified by the detrimental effects either in temporal dimension (indicated by a residual queue) or in spatial dimension (indicated by a spillover). With the information of OSI, a maximum flow based signal control model is formulated to manage oversaturation. The proposed model can effectively and efficiently manage oversaturated routes without the requirement on time-dependent traffic demand as model inputs. Instead of dealing with hundreds of decision variables and constraints, the control model can be solved by a simple Forward-Backward Procedure (FBP), which makes the practical use of the model much more promising.

3) **It builds a new integrated control model to reduce network congestion through diversion control.** The model can diagnose incidents, measure performance during the occurrence of incidents and suggest optimized signal control strategies in real time. Comparing with the previous control models, the proposed one specifically considers the impact of the diversion traffic to diverting route, especially for signalized arterials, so the potential congestion caused by

diversion traffic can be reduced or eliminated by proper adjustment of signal timings. It does not have the requirement on time-dependent traffic demand information as model inputs. It has very low computation burden and is suitable for on-line applications.

1.3 Dissertation Organization

The rest of this dissertation is organized as follows. Chapter 2 briefly introduces the SMART-Signal system, which provides all the required data in this research. Chapter 3 presents an arterial offset optimization model based on the high-resolution traffic signal data. In Chapter 4, a maximum flow based control model is developed to manage oversaturated arterials. An integrated control model for managing network congestion is introduced in Chapter 5. Finally, the conclusion and future research are included in Chapter 6.

2. The SMART Signal System

2.1 Background

Although measuring and archiving freeway traffic performance has become a norm for many transportation agencies for years, similar approaches for urban signalized arterials do not exist. In practice, arterial performance measures are usually evaluated in an ad hoc manner, involving manual data collection and data processing, which is time-consuming and costly. Operational data from traffic signal systems (such as vehicle-detector actuations and signal phase changes) are neither stored nor analyzed, which prevents the proactive management of arterial streets.

The SMART-Signal (**S**ystematic **M**onitoring of **A**rterial **R**oad **T**raffic **Signal**) system, developed at the University of Minnesota, fills in this gap. It automatically collects event-based high-resolution traffic signal data from multiple intersections and generates real-time arterial performance measures including intersection queue length, delay, level of service, arterial travel time, link delay and etc. Archived traffic performance measures are also used for parameter fine-tuning for traffic signal control systems. The system is now installed in 13 intersections on TH 13 (in Burnsville, MN) and 10 intersections on TH55 (in Golden Valley, MN), 22 intersections on TH 7 (in Minnetonka, MN), 25 intersections on TH 65(in Blaine, MN), 21 intersection on TH 61(in White Bear Lake, MN), 5 intersections on TH 10 (in Anoka, MN) and 6 intersections on Orange Grove Boulevard in Pasadena, CA. The system has laid the groundwork for better traffic signal operations and opens up entirely new opportunities for managing traffic on congested roads.

2.2 Data Collection Units

The SMART-Signal Data Collection Units (DCUs) are utilized to collect event-based data (e.g. detector actuations and signal status changes) from signalized intersections. Until now, two types of data collection units (see Figure 2-1) have been developed for different control cabinets. For the TS-1, 170 and 2070 type cabinet, the DCU detects the

voltage changes on the back-panel, which indicate different types of traffic events in field; For the TS-2 type cabinets, the DCU captures the event codes directly from the bus line through the SDLC port. All time-stamped signal status changes and vehicle-detector actuations can be captured by the DCUs and transmitted back to the data server in real-time (see Figure 2-2) and stored in the database of the traffic management center.

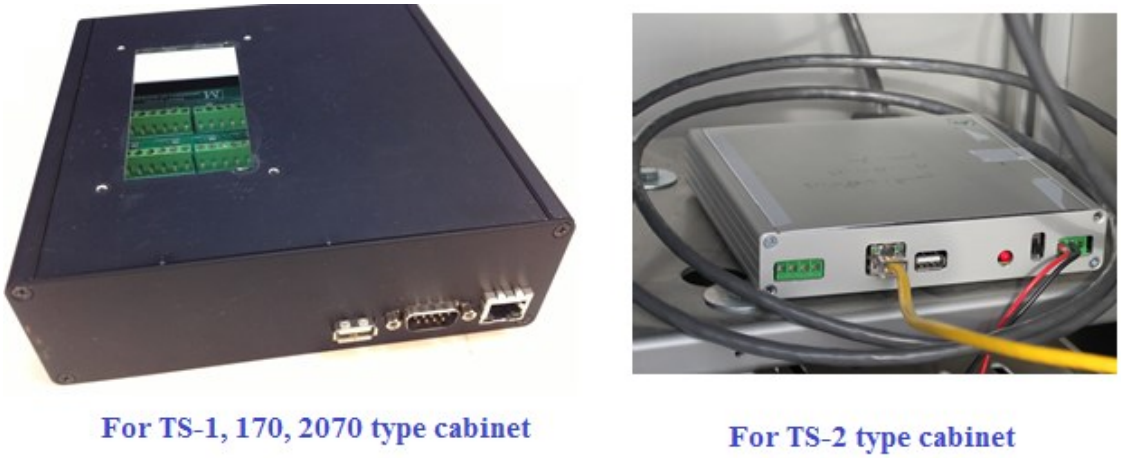


Figure 2-1 SMART-Signal data collection units

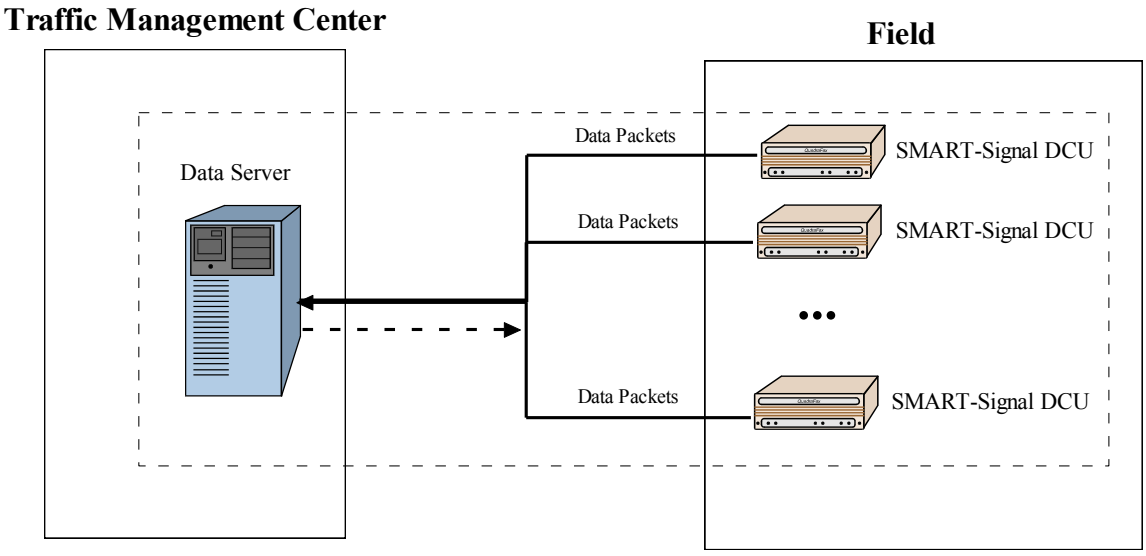


Figure 2-2 SMART-Signal data transmission

Two sample data tables for signal status changes and detector actuations are shown in Figure 2-3. The time stamp, phase/detector ID and time duration are recorded for each event, based on which a series of performance measures can be derived.

	ID	TimeStamp	Status	PhaseID	TimeDuration
1	1	130501000547000	Y	2	5.000
2	2	130501000547000	Y	6	5.000
3	3	130501000553500	G	4	9.900
4	4	130501000603400	Y	4	3.500
5	5	130501000552000	R	2	16.900
6	6	130501000552000	R	6	16.900
7	7	130501000608900	G	2	109.700

For signal status changes

	ID	TimeStamp	DetectorID	TimeDuration
1	8521	130401162230100	16	4.700
2	8522	130401162235700	16	0.800
3	8523	130401162238800	16	1.200
4	8524	130401162235400	6	5.900
5	8525	130401162242400	16	1.400
6	8526	130401162244300	2	0.300
7	8527	130401162241800	6	7.500

For detector actuations

Figure 2-3 SMART-Signal data tables

2.3 Performance Measures

Based on the archived event-based data, a set of arterial performance measures can be estimated. Two innovative models have been developed to estimate the two most important measures for signalized arterials, i.e. intersection queue length (Liu et al., 2009) and arterial travel time (Liu and Ma, 2009), based on which other performance measures such as average delay, level of service and link delay can be easily derived. For a more complete list of performance measures that the SMART-SIGNAL system provides, please see http:// dotapp7.dot.state.mn.us/smartsignal.

The collected high-resolution event-based data and derived performance measures by the SMART-Signal system are very useful and valuable. As you will see in the rest of this dissertation, they can be used to optimize offsets for vehicle-actuated coordinated traffic signals, manage oversaturated arterials and control integrated corridors.

3. Arterial Offset Optimization Using Archived High-resolution Traffic Signal Data

3.1 Background

Offset settings are critical to the performance of traffic signal system operated in a coordinated mode. Over the years many computerized programs have been developed for signal offset optimization, using average traffic volumes at intersections and average travel times between intersections. Based on the objectives of the optimization process, these programs can be divided into two groups. One group aims to maximize the progression bandwidth along arterials, such as MAXBAND (Little et al., 1981) and PASSER (Messer et al., 1974), and the other group attempts to minimize system delay and number of stops, such as TRANSYT (Robertson, 1967) and Synchro (Trafficware, 2001). In more detail, MAXBAND maximizes weighted combination of bandwidths given cycle length and splits. It allows users to input a queue clearance time for each approach, but the queuing process is not considered inside the model. PASSER maximizes the bandwidth efficiency based on pre-calculated splits. Traffic performance measures are estimated using deterministic models. TRANSYT is a macroscopic deterministic optimization model which minimizes a performance index, defined as a weighted combination of system delay and stops. It models traffic in platoons using a platoon dispersion algorithm and considers queuing process by estimating arrival and discharging flow second by second. MAXBAND, PASSER and TRANSYT can all be considered as deterministic models because they only use average traffic volumes as optimization inputs. Traffic flow fluctuations are not considered in these models. As argued by Heydecker (1987), the degree of variability of traffic flows and signals has a significant impact on the outcome of signal optimization, in that using the average flows may incur considerable additional delay, compared with the timing plan obtained by taking this variability into account. In order to accommodate traffic variations, Synchro optimizes signal timing by averaging results of five volume scenarios (10, 30, 50, 70 and

90 percentiles), which assumes the user-input volumes are the means of random variables with Poisson distribution. However, these scenarios are arbitrarily assumed and they may not be consistent with actual traffic conditions in the field.

As indicated by Abbas et al. (2001), existing offset optimization models for vehicle-actuated coordinated traffic signals fails to address two important issues: (1) the early return to green problem because of the traffic flow variations on non-coordinated approaches, (2) the variation of waiting queue lengths at coordinated directions. Over the years, a number of research studies on offset optimization have attempted to address these two problems from different perspectives. To mitigate the "early-return-to-green" problem, Skabardonis (1996) developed several alternative procedures to modify controller yield points and phase force-offs using either the average spare green times or the average duration of the green times for the actuated phases. Shoup and Bullock (1999) developed an offline procedure to optimize offsets, assuming the availability of travel times between intersections from vehicle re-identification technologies. Abbas et al. (2001) developed a real-time offset transitioning algorithm for coordinated traffic signals. The algorithm uses a greedy approach to search for the optimal offset by moving the green window to include more vehicles that can pass during green time. Gettman et al. (2007) proposed a real-time control algorithm to adjust intersection offsets in a coordinated traffic signal system, based on the signal phase and detector data from the last several cycles. Using the archived signal status data, Yin et al. (2007) presented an offline offset refiner, which addresses the problem of uncertain green starts and ends in the determination of offsets. Later, Zhang and Yin (2008) extended the study and developed a robust approach to the synchronization problem of actuated signals along corridors. More recently, based on high resolution traffic signal data, Day et al. (2010) introduced the Purdue Coordination Diagram (PCD) to assess arterial coordination and Day et al. (2011) evaluated the performance of different offset optimization objective functions. However, none of the previous studies have addressed the two problems identified above simultaneously, which could lead to suboptimal solutions. These two

problems are intertwined together, as traffic demand variations will lead to uncertain green start times, and uncertain green start times will result in stochastic length of waiting queues even with deterministic demands. The two problems, therefore, need to be considered simultaneously.

In this chapter, a data-driven approach is developed to optimize offsets for vehicle-actuated coordinated traffic signals, using the massive amount of signal status and vehicle actuation data collected from field. The model is intended to serve as an off-line offset fine-tuning tool, which optimizes the offsets after monitoring the traffic condition in the field for a certain period of time. The proposed approach, for the first time, takes the two problems of vehicle-actuated traffic signal coordination into consideration simultaneously. To account for the early return to green problem, for the coordinated phase, the concept of conditional probability of the green start times is introduced, which can be obtained from the archived high-resolution traffic signal data. To handle the uncertainty of intersection queue length, a scenario-based optimization approach is adopted, which uses a series of traffic demand scenarios obtained from the archived data as the model input. The objective of this model is to minimize total delay for the main coordinated direction and at the same time considers the performance of the opposite direction. We test the performance of the optimized offsets in the field and the result shows that the proposed model can reduce travel delay of coordinated direction without compromising the performance of the opposite direction. In practice, we envision that the optimization program can be utilized periodically (for example, every a few weeks) to optimize system performance using the archived data during that period.

In section 3.2, we present the high-resolution traffic signal data and introduce two important concepts: the conditional distribution of green start times and the traffic demand scenarios. In section 3.3, the scenario-based offset optimization model is presented. The solution method and optimization results are given in section 3.4. Results

from the field test are described in section 3.5. The summary of this chapter is included in section 3.6.

3.2 High-resolution Traffic Signal Data

The proposed offset optimization model is based on the traffic data collected by the SMART-SIGNAL system. As discussed in Chapter 2, the system is an integrated data collection, storage, and analysis system which can continuously collect and archive high resolution event-based traffic signal data including every vehicle-detector actuation and every signal phase change. We should note that every vehicle-detector actuation event is time-stamped, and this is in contrast to legacy logging systems that would bin detector data on a minute or greater basis.

Six intersections along the TH55 in Golden Valley, MN were chosen as the study site, as shown in Figure 3-4. This corridor has a total length of 1.83 miles with a speed limit of 55 MPH. The six intersections are operated in a vehicle-actuated coordinated mode during weekday's morning peak (7:00AM ~ 9:00 AM) and afternoon peak (3:30 PM ~ 5:30 PM), with a cycle length of 180 seconds. The coordination favors east bound (phase 2) traffic in the morning peak and west bound (phase 6) traffic in the afternoon. Ten weekdays' data during morning peak hours from 6/15/2009 to 6/26/2009 were used in this research.

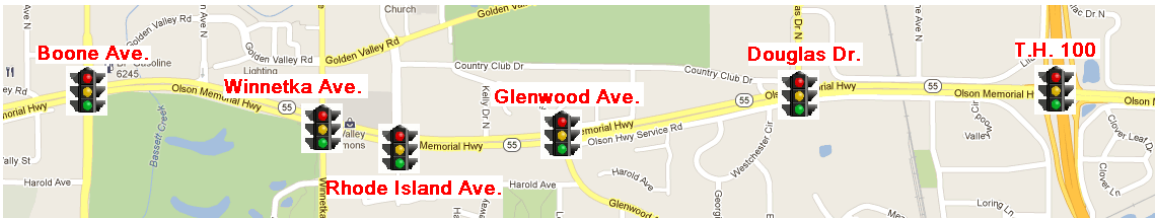


Figure 3-4. Installation of the SMART-SIGNAL system on TH55, Golden Valley, MN (Source: Google Maps)

3.2.1 Conditional Distribution of Green Start Times

As mentioned above, the SMART-SIGNAL system archives every signal change and detector actuation at signalized intersections. Figure 3-5 presents an example distribution of green start time of the coordinated phase (Phase 2) in the local controller clock at the intersection of TH55 & Boone Ave. Figure 3-6 shows the corresponding dual-ring control diagram, which has 8 phases. Based on the settings, phase 2 and phase 6 are the designated coordinated phases and a cycle starts at the time when phases 3 and 8 begin. Under different traffic conditions, the non-coordinated phases may be terminated due to gap-out or max-out, or at its force-off point. It can even be skipped if there is no vehicle actuation for a particular phase. Any unused green time from the non-coordinated phases would be added to the coordinated phases (i.e. phase 2 and phase 6), which could make the coordinated phases start earlier. This is the so-called "early return to green" problem. In this case, the duration of phases 3 and 4 and phases 7 and 8 would determine the green start time point of phase 2, as can be seen from Figure 3-6. The minimum split length for phase 3 and phase 4 are 12 seconds and 16 seconds respectively and their maximum splits are 15 seconds and 26 seconds. One pedestrian phase is associated with phase 4, which has a "Walking" time of 7 seconds and a flashing "Don't Walk" time of 30 seconds. Since the assigned pedestrian phase is longer than the duration of associated phase 4, if there is a pedestrian call, it will extend the phase 4 green beyond its original maximum. As a result, the green start time distribution in Figure 3-5 is clearly divided into three groups, as indicated by the numbers in red. In group 1, either phase 3 or phase 4 is skipped, which makes phase 2 green start time smaller than the summation of two minimum split lengths (12 + 16 = 28 seconds); In group 2, both phases are active in the cycle and the total spit length varies from its minimum to maximum (15 + 26 = 41 seconds); In part 3, there is at least one pedestrian call in the cycle which makes coordinated phase 2 starts later than usual.

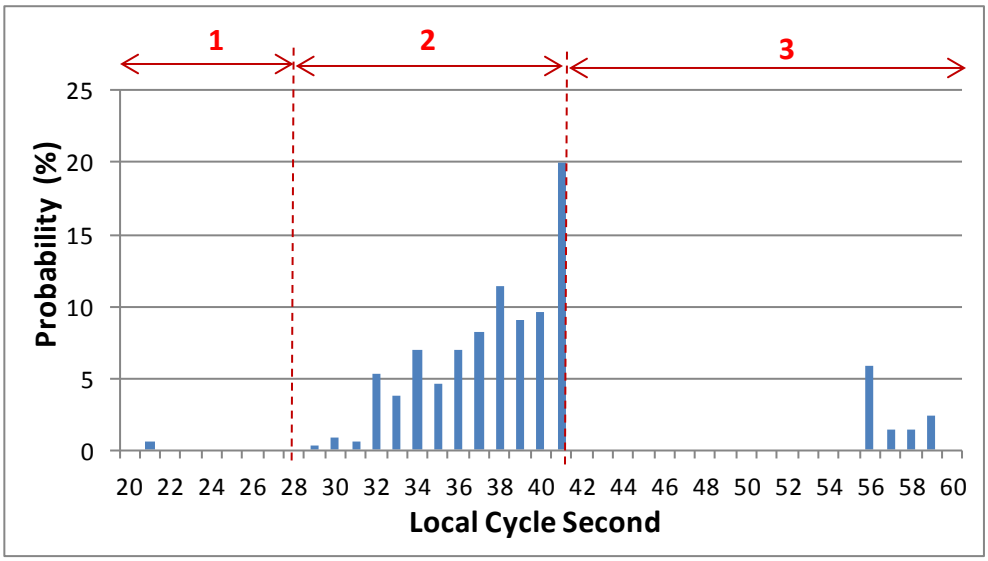


Figure 3-5 Distribution of green start time of Phase 2 (Coordinated phase) in local clock, TH 55 & Boone Ave, 7:00 AM ~ 9:00 AM, 6/15/2009 ~ 6/26/2009 10 weekdays

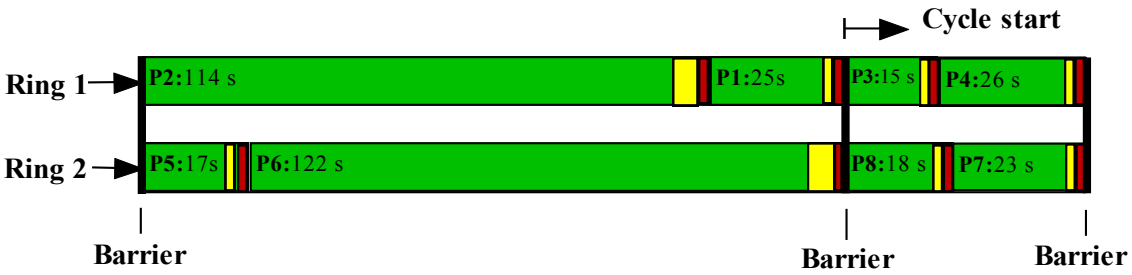


Figure 3-6 Dual-ring control diagram of intersection TH 55 & Boone Ave during weekday AM peaks

As shown above, at vehicle-actuated intersections, signal timings vary from cycle to cycle under different actuation patterns and the "early return to green" problem of coordinated phases exist due to various reasons. The green start time of each coordinated phase is mainly dependent on the vehicle actuations of non-coordinated phases. To represent this relationship, an associated number of vehicle actuations, $\rho_{n,i}$ for coordinated phase i of intersection n, is introduced. Without loss of generality, a four-leg and 8-phase intersection is taken as an example. Intersection layout and corresponding dual-ring

control diagram are shown in Figure 3-7. The numbers in the layout indicate different phases (phase 2 and phase 6 are coordinated phases here) and the red arrows indicate different moving directions. According to the dual ring control logic, different rings can only go across the "barriers" at the same time. Specifically, phase 3 and phase 8 should start at the same time and phase 4 and phase 7 should end at the same time. As a result, the non-coordinated phase in one ring with less vehicle actuations, which is supposed to end earlier, may be extended by the phases in the other ring with more vehicle actuations.

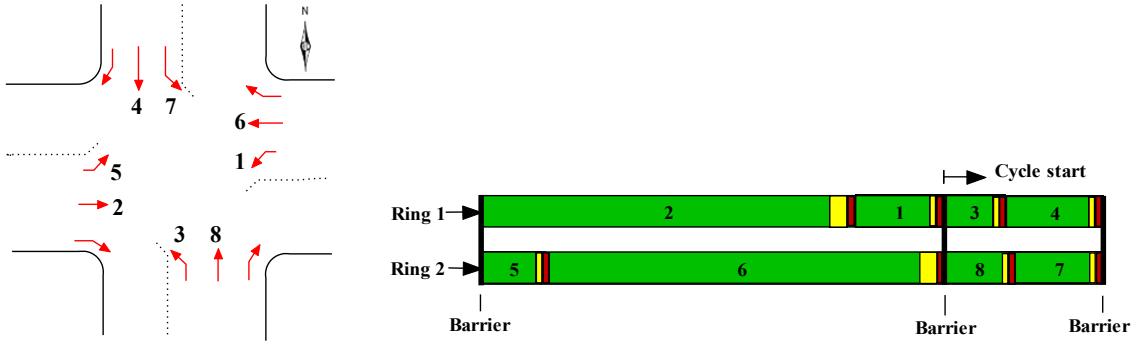


Figure 3-7 Dual Ring Control diagram of an example intersection

Because the associated number of actuations $\rho_{n,i}$ in a cycle is dependent on the input volumes for phases preceding the coordinated phase, the calculation of $\rho_{n,i}$ can be shown in Equation(3-1), where $V_{n,i}$ is the cycle input volume for phase i of intersection n. It states that the associated number of vehicle actuations of phase 2 is equal to the larger values between $V_{n,3}+V_{n,4}$ and $V_{n,7}+V_{n,8}$. For phase 6, the associated number of vehicle actuations is equal to the larger values between $V_{n,3}+V_{n,4}$ and $V_{n,7}+V_{n,8}$ plus the number of actuations of phase 5 ($V_{n,5}$). Note that phase 1 is not included when calculating the associated number of vehicle actuations for phase 2, because the lagging phase after the coordinated phase usually has a fixed length of duration.

$$\rho_{n,2} = \max(V_{n,3} + V_{n,4}, V_{n,7} + V_{n,8})$$
$$\rho_{n,6} = \max(V_{n,3} + V_{n,4}, V_{n,7} + V_{n,8}) + V_{n,5} \quad (3\text{-}1)$$

It is intuitive that the larger the $\rho_{n,i}$ is, the later the green phase of the coordinated phase i tends to start. When we consider different arrival and actuation profiles of non-coordinated phases, the green start time of coordinated phase i can be different even if $\rho_{n,i}$ is the same. Therefore, for coordinated phase i, given its associated number of actuations $\rho_{n,i}$, we can obtain the distribution of green start time $G_{n,i}$. Note that, the introduction of the conditional distribution $P\left(G_{n,i} \mid \rho_{n,i}\right)$ is specifically designed to accommodate the "early return to green" problem for the coordinated phases. Even if the input volumes are the same, actual signal timings may still vary, and the variation is consistent with field results since all the distributions are derived from field data.

For example, Figure 3-8 shows the green start times of coordinated phase 2 at the intersection TH 55 & Boone Ave according to the associated number of actuations $\rho_{1,2}$. The data points are clearly divided into three groups, as discussed before. One can see that, within each group, as the associated number of actuations $\rho_{1,2}$ increases, the green start time of phase 2 tends to be later. Therefore, different values of associated number of vehicle actuations will lead to different distributions of green start time. The two plots in Figure 3-9 show the conditional distributions when $\rho_{1,2} = 6$ and $\rho_{1,2} = 9$ respectively. We can see that there are two different distribution patterns. When $\rho_{1,2} = 6$, the green start time are almost evenly distributed between 35 and 41 seconds; however, when $\rho_{1,2} = 9$, 45% of the time, phase 2 green would start at 41 seconds. Similarly, conditional distributions can be generated for phase 6 and also for other intersections.

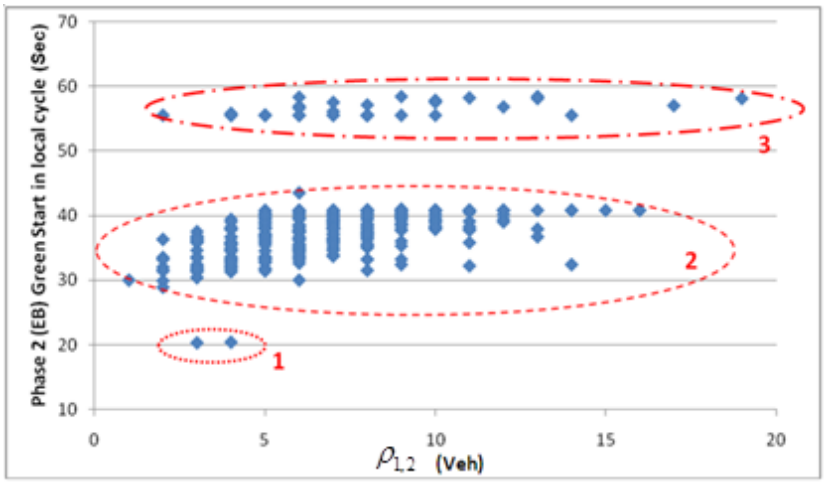


Figure 3-8 Phase 2 (Eastbound through) green start time vs. associated number of actuations $\rho_{1,2}$, TH 55 & Boone Ave, 6/15/2009 ~ 6/26/2009 10 weekdays

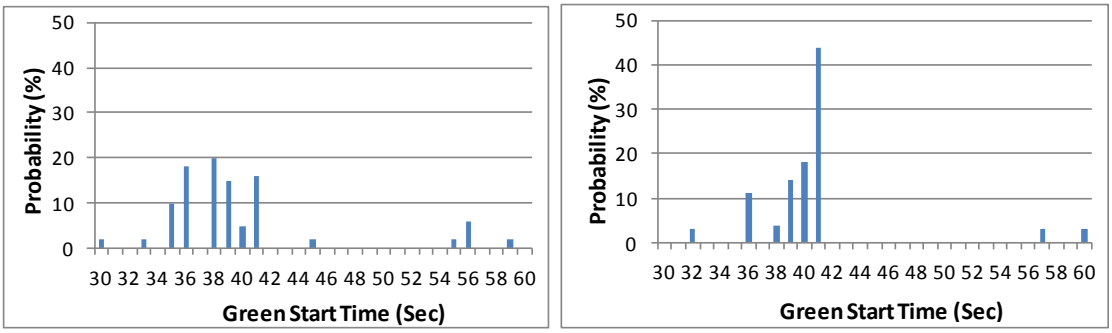


Figure 3-9 Conditional distributions of Phase 2 (EB) green start time $P(G_{1,2} \mid \rho_{1,2})$, (Left) $\rho_{1,2}$=6 (Right) $\rho_{1,2}$=9 ,TH 55 & Boone Ave, 6/15/2009 ~ 6/26/2009 10 weekdays

3.2.2 Traffic Scenarios

Based on the high-resolution traffic signal data, traffic demands at the boundary of the arterial and turning percentages at each intersection can also be obtained. As shown in Figure 3-10, input volume of direction d at intersection n is denoted by $U_{n,d}$ ($n = 1, 2, \ldots, d =$ *Eastbound, Westbound, Northbound, or Southbound*) and

corresponding turning percentages are denoted by $\varphi_{n,d}^{x}$ (x =*Left, Right, Through*). Without loss of generality, we assume the arterial is coordinated in the east-west direction. To deal with the uncertainty of traffic demand, a set of traffic scenarios $S = \{1, 2, ..., J\}$ is introduced. Each scenario $j \in S$ is a vector of boundary input volumes and turning percentages, i.e. $\left(U_{1,EB}^{(j)}, U_{1,SB}^{(j)}, ..., \varphi_{1,EB}^{L(j)}, \varphi_{1,EB}^{R(j)}, ...\right)^T$.

In general, the more scenarios we consider, the more robust the optimization results will be. But that will inevitably increase the complexity of the problem. According to Mulvey et al. (1995), a relatively small number of scenarios can still generate near-optimal results. In practice, the number of scenarios considered in the optimization model is constrained by the availability of dataset, computational capability and etc, so users can choose the most suitable number of scenarios accordingly. As an example, 10 weekdays' (6/15/2009 ~ 6/26/2009) morning peak (7:00 AM ~ 9:00 AM) data are utilized and every 15 minutes' interval is considered as one traffic scenario. To check the sufficiency of the data set, we plotted the cumulative distributions of the 15 minutes' input volumes for the eastbound and westbound directions on TH55. The distributions are generated using 5 days, 10 days and 20 days of AM peak data, and the results are shown in Figure 3-11. As one can see that the distribution lines are very close to each other in each figure. The Kolmogorov–Smirnov test has been conducted to test the null hypothesis that the 15 minutes' input volumes from the 10 days and 20 days of data have the same distribution. For eastbound input volumes, the *p* value is 0.9147; for westbound input volumes, the *p* value is 0.9821. Both high *p* values returned from the K-S test indicate the similarity of the two samples. Therefore, the 10 weekdays' morning peak data set is large enough to estimate the field distribution. The test was conducted to other intersections and approaches, and the same conclusion was drawn.

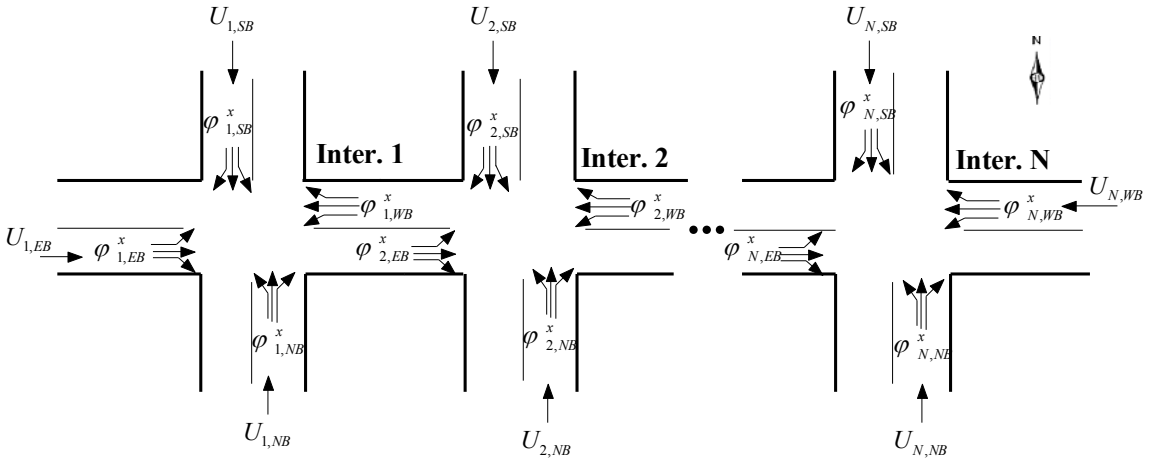


Figure 3-10 Boundary input volumes and turning percentages

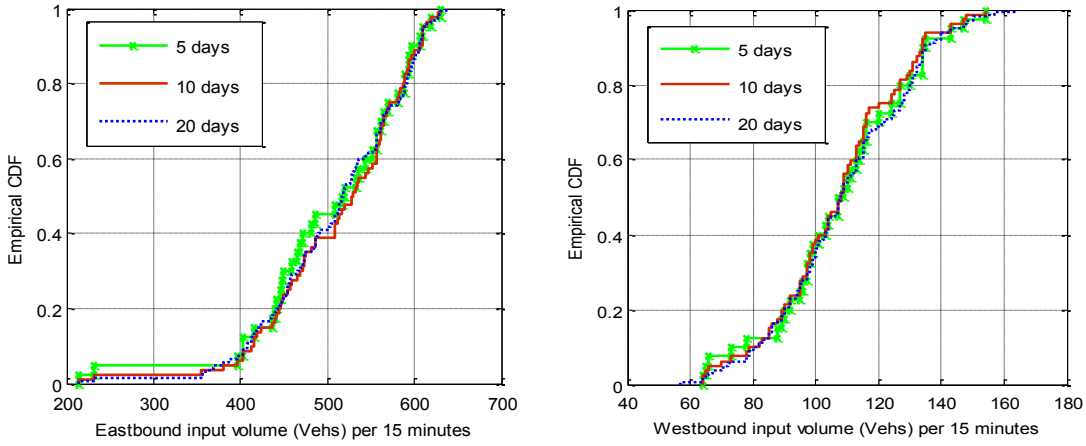


Figure 3-11 Empirical cumulative distributions for eastbound and westbound input volume

In total there are 80 traffic scenarios and within each scenario $j \in S$, the boundary input volumes $U_{n,d}^{(j)}$ and turning percentages $\varphi_{n,d}^{x\ (j)}$ are assumed to be stable and the values are set as the averages over all cycles within the 15-minute interval. Given a scenario $j \in S$, the cycle input volumes for intermediate links can be calculated by (3-2) and (3-3). It is easy to find out that eastbound cycle input volume of intersection $n+1$ is the summation of volumes of three movements at intersection n (eastbound through, southbound left and

northbound right), as shown in (3-2). Similarly, westbound cycle input volume of intersection *n-1* is calculated by (3-3). By applying (3-2) and (3-3) to each intersection, the cycle volume of each movement becomes known.

$$U_{n+1,EB}^{(j)} = U_{n,EB}^{(j)} \times \varphi_{n,EB}^{T\,(j)} + U_{n,SB}^{(j)} \times \varphi_{n,SB}^{L\,(j)} + U_{n,NB}^{(j)} \times \varphi_{n,NB}^{R\,(j)} \tag{3-2}$$

$$U_{n-1,WB}^{(j)} = U_{n,WB}^{(j)} \times \varphi_{n,WB}^{T\,(j)} + U_{n,SB}^{(j)} \times \varphi_{n,SB}^{R\,(j)} + U_{n,NB}^{(j)} \times \varphi_{n,NB}^{L\,(j)} \tag{3-3}$$

3.3 Model Formulation

3.3.1 Scenario-based Offset Optimization

To account for the stochastic nature of field traffic, a data-driven optimization model is developed to generate arterial offsets. As discussed in the previous section, based on the high-resolution traffic signal data, the conditional distributions of green start times P and traffic scenarios S can be derived, which will be the model inputs. In general, the optimization model can be illustrated by (3-4), where the objective is to minimize the total travel delay D of coordinated directions. O is the vector of intersection offsets (decision variables) and Ω is the feasible set of O.

$$\begin{aligned} &\min D(O,P,S) \\ &s.t.\ \ O \in \Omega \end{aligned} \tag{3-4}$$

Given a traffic scenario $j \in S$, since the cycle input volume for each movement can be calculated, the associated number of vehicle actuations, denoted by $\rho_{n,i}^{(j)}$, can be generated for the coordinated phases of intersection *n*. Then, given a feasible offset $O \in \Omega$, the expectation of the total control delay for coordinated phase i of intersection *n* under scenario j, denoted by $E\left(D_{n,i}^{(j)}(O)\right)$, can be calculated by (3-5). Here O represents a vector of offsets $[o_1, o_2, ..., o_N]$ and o_n is defined as the time difference

from the start of red of the coordinated phase at intersection *n* to that of the master intersection. The calculation of control delay $D_{n,i}^{(j)}(O)$ under scenario *j* and given signal offsets O will be discussed in the next section. Here *x* and *y* represent any feasible green start times for phase *i* at intersections *n* and *n+1*. In other words, they represent any feasible realization of random variable $G_{n,i}$. Note that, the conditional distribution of green start time for the coordinated phase at intersection n is assumed to be independent with that at intersection $n+1$. This assumption is reasonable because the green start time of coordinated phase at each intersection is determined by the corresponding vehicle actuations of non-coordinated phases and it is reasonable to assume they are independent.

$$E\left(D_i^{(j)}(O)\right) = E\left(\sum_{n=1}^{N-1} D_{n,i}^{(j)}(O)\right) = \sum_{n=1}^{N-1} E\left(D_{n,i}^{(j)}(O)\right)$$
$$= \sum_n \sum_x \sum_y D_{n,i}^{(j)}(O) P\left(G_{n,i} = x \mid \rho_{n,i} = \rho_{n,i}^{(j)}\right) P\left(G_{n+1,i} = y \mid \rho_{n+1,i} = \rho_{n+1,i}^{(j)}\right) \quad (3\text{-}5)$$

An ideal solution for arterial coordination would optimize the performance of two major directions at the same time. However, in most cases, the two directions cannot achieve their best performance simultaneously and there is always a trade-off between the two. Therefore, a weighting term $\alpha \in [0,1]$ is introduced to represent the importance of the minor coordinated direction and the weighted delay for both directions is calculated by (3-6), where $D_i^{(j)}(O)$ and $D_{i'}^{(j)}(O)$ are the travel delays for major and minor coordinated directions respectively.

$$D^{(j)}(O) = (1-\alpha) E\left(D_i^{(j)}(O)\right) + \alpha E\left(D_{i'}^{(j)}(O)\right) \quad (3\text{-}6)$$

Given an intersection offset setting O, we can calculate the corresponding arterial delay series $D^{(1)}(O), D^{(2)}(O), ..., D^{(J)}(O)$ for different scenarios. Let $D_{(1)}(O), D_{(2)}(O), ..., D_{(J)}(O)$ be the descending order of the series, i.e., $D_{(1)}(O) \geq D_{(2)}(O) \geq, ..., \geq D_{(J)}(O)$. Obviously, the larger the value of D is, the worse the arterial performance will be. In practice, motorists and engineers may have different opinions about these scenarios and therefore a trimming factor $\beta \in [0,1]$ is introduced. The value of β represents the fraction of scenarios that will not be considered and the number of scenarios left after trimming is $J_\beta = \lfloor J(1-\beta) + \beta \rfloor$. If $\beta = 0$, there is no trimming ($J_\beta = J$) and all the scenarios are considered; However, if $\beta = 1$, $J_\beta = 1$, only the worst scenario $D_{(1)}(O)$ will be considered. Therefore, the measure of arterial delay given intersection offset setting O can be formulated as $\frac{1}{J_\beta}\sum_{k=1}^{J_\beta} D_{(k)}(O)$ and $D_{(k)}(O)$ is the k-th largest delay value. The procedure of data processing is illustrated in the flow chart shown in Figure 3-12, which summarizes how to process the data from the archived traffic signal data set and calculate the total delay given a set of intersection offsets. The general form of the scenario-based offset optimization model is shown in (3-7) and the deterministic delay calculation model will be introduced in the next section. Note that, under transformed time coordinates (i.e. shifted according to the free flow travel time from the first intersection), intersection offsets vary from $-c/2$ to $c/2$, where c is the common cycle length.

$$\begin{aligned} &\min \frac{1}{J_\beta}\sum_{k=1}^{J_\beta} D_{(k)}(O) \\ &s.t.\ O \in \Omega, \\ &\Omega = \left\{ [o_1, o_2, ..., o_N] : -\frac{c}{2} \leq o_n \leq \frac{c}{2}, \forall n \right\} \end{aligned} \tag{3-7}$$

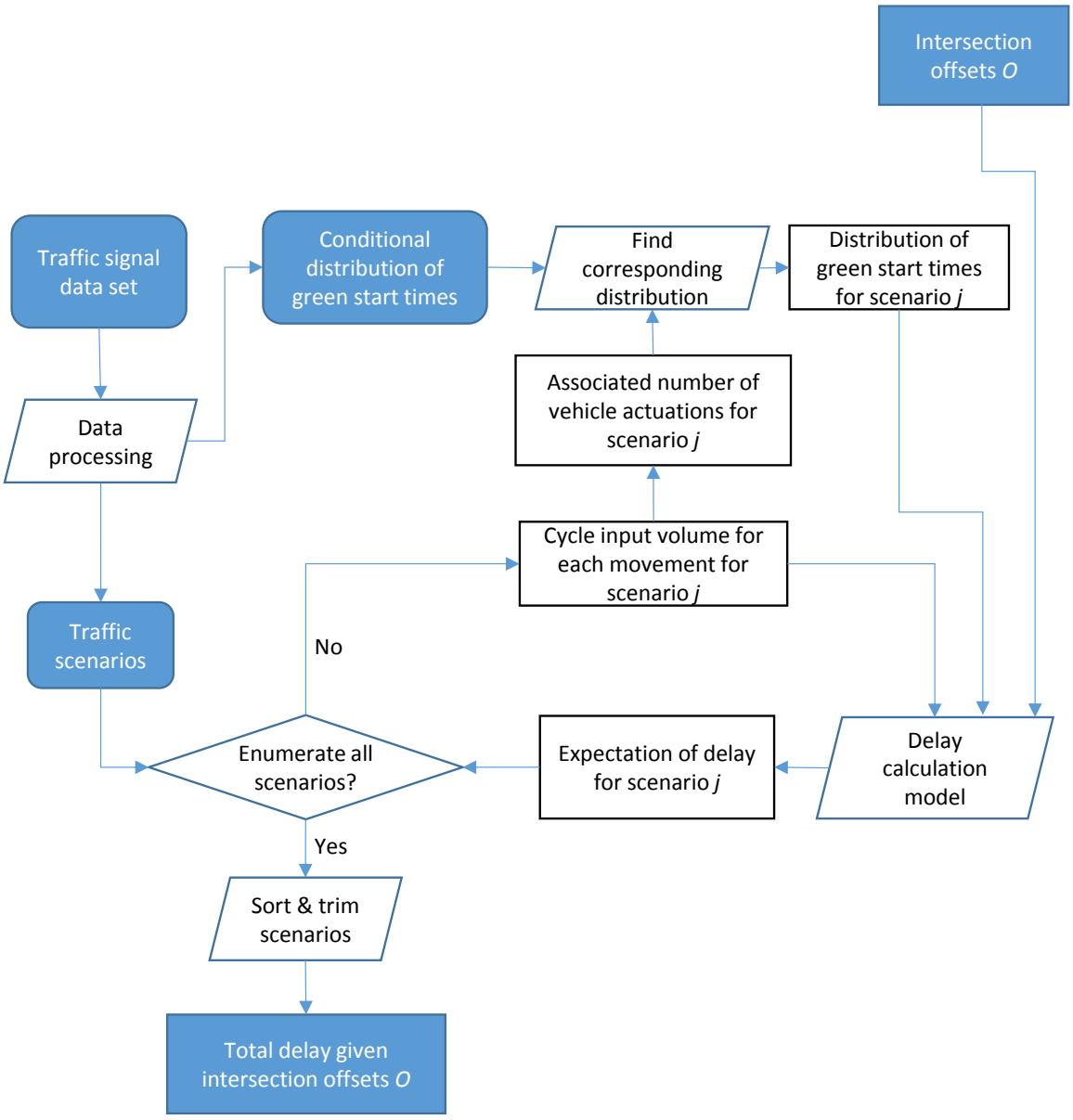


Figure 3-12 Data processing flow chart

3.3.2 Delay Calculation

Traditionally, the delay at an intersection is approximated by two terms: uniform (deterministic) delay and random delay, as described in the Highway Capacity Manual (2010). The uniform delay can be calculated by assuming an average arrival rate, while

the random delay component can be calculated based on the probability distribution of time gaps between successive vehicles, given a fixed average arrival rate. However, in reality, there are actually two other types of randomness that need to be considered in the calculation of delay at a signalized intersection: (1) traffic demand variation among different scenarios, in order words, average arrival rates are different across different demand scenarios; (2) the randomness of green start times because of the early-return-to-green problem for coordinated actuated signals.

In this section, a basic delay calculation model and an extended model will be introduced. The basic model assumes that vehicles are uniformly discharged during red and green periods of the coordinated phase. However, we adopt a scenario-based approach to handle the variation of field traffic demand and use the conditional distribution of the green start times for the coordinated phase (discussed in Section 3.2.1) to account for the early return to green problem. For the extended model, we further relax the assumption of uniform arrival rates during green and red periods in a cycle and utilizes the discharging profiles of each cycle at boundary links to derive intersection delay. Because of the availability of high-resolution loop detector data, we know exact traffic arrivals (i.e., time gap between consecutive vehicles can be measured) at boundary links for every second, i.e., the arrival profile at every second can be built using the high-resolution data. Therefore we can actually calculate delay deterministically without consideration of randomness of vehicle's inter-arrival times.

Given intersection signal timings (i.e., offsets and green start times for coordinated phases) and traffic scenario (i.e., traffic demand or discharging profiles), the calculation of delays for coordinated directions becomes a deterministic problem. Since the red light durations of coordinated phases have direct association with the calculation of travel delay for coordinated directions, for the ease of presentation, we use red time durations $r_{n,i}$ in the delay calculation instead of green start times $G_{n,i}$ because they are the same

numerically. We further omit subscript *i* and use r_n to represent red light duration, when the indication of phase is clear.

(1) Basic Model

The calculation for one coordinated direction is first considered since the calculation for the opposite direction can follow the same way. In the following, we take the eastbound (a coordinated direction) as an example. At each intersection *n*, cycle input volumes to downstream can be separated into two parts. One is discharged from side streets when coordinated phase is red, denoted by W_n^s; the other is discharged from main street when coordinated phase is green, denoted by W_n^m. For example, for eastbound traffic, W_n^s is equal to $U_{n,SB} \times \varphi_{n,SB}^L + U_{n,NB} \times \varphi_{n,NB}^R$ and W_n^m is equal to $N_{n,EB} \times \varphi_{n,EB}^T$. Assume vehicles are uniformly discharged during red and green periods of the coordinated phase and they are evenly distributed at downstream lanes, the arrival flow rate per lane to downstream intersection in these two periods can be calculated by (3-8), where q_n^s and q_n^m (Veh./Sec./Ln.) are the arrival flow rate per lane to downstream intersection formed by side street traffic (during coordinated red) and main street traffic (during coordinated green) of intersection *n* respectively, and z_{n+1} is number of lanes at downstream link.

$$q_n^s = \frac{W_n^s \varphi_{n+1,EB}^T}{r_n z_{n+1}}, \quad q_n^m = \frac{W_n^m \varphi_{n+1,EB}^T}{(c - r_n) z_{n+1}} \tag{3-8}$$

In the delay calculation model, time coordinate of each intersection is shifted according to the free flow travel time from the first intersection. Similar treatment for delay calculation can date back to Bavarez and Newell (1967). Vehicle trajectory then becomes a vertical line in the time-space diagram if there is no stop or delay (see Figure 3-13 and Figure 3-14). Assume the offset reference point is the green end (or red start) of

corresponding coordinated phase. Two variables d_n^c and d_n^g are introduced to represent the time difference of cycle starts and green starts between intersection n and downstream intersection $n+1$ under the transformed time coordinates. The definition is shown in (3-9), where o_n is the offset value and r_n is the red duration time of coordinated phase at intersection n. Based on the sign of d_n^c, the delay calculation process is divided into two cases.

$$d_n^c = o_{n+1} - o_n$$
$$d_n^g = o_{n+1} + r_{n+1} - o_n - r_n \quad (3\text{-}9)$$

a) $d_n^c \geq 0$

In this case the upstream red light for the coordinated phase starts earlier than that of the downstream intersection (See Figure 3-13). Depending on the start time of the downstream green light, vehicles discharged from upstream side streets with an arrival rate q_n^s may cause a queue at the downstream intersection; After the upstream intersection turns green, downstream intersection will have arrival vehicles from upstream coordinated phase with a rate of q_n^m. Under different values of d_n^g and arrival flows, this portion of vehicles may or may not cause a queue at downstream. For example, if $d_n^g \geq 0$, as shown in Figure 3-13, upstream green starts earlier than downstream and part of the upstream main street flow will definitely cause a queue at downstream; if $d_n^g < 0$, downstream green starts earlier than upstream, so the queue may be cleared before the arrival of upstream main street vehicles. If t_n^q is used to represent the portion of green at intersection *n* within which discharged vehicles will cause a queue at downstream and q_{n+1}^d(Veh/sec) is the discharging rate of intersection $n+1$, the following flow conservation equation holds.

$$(r_n - d_n^c)q_n^s + t_n^q q_n^m = (t_n^q - d_n^g)q_{n+1}^d \tag{3-10}$$

t_n^q can be calculated by solving equation (3-10). $t_n^q = 0$ when upstream main street flow does not cause a queue at downstream. Since t_n^q cannot be negative,

$$t_n^q = \max\left\{\frac{(r_n - d_n^c)q_n^s + d_n^g q_{n+1}^d}{q_{n+1}^d - q_n^m}, 0\right\} \tag{3-11}$$

The first vehicle from coordinated phase of intersection n which has no delay at downstream will have the trajectory of CC' in Figure 3-13. Note that h represents the jammed space headway in the figure. The area of the shadowed triangle ΔDEF divided by h represents the total delay (Veh.Sec) of all the vehicles discharged from coordinated phase of intersection n in one cycle. The delay D_n can be calculated by (3-12). Note that the area of the triangle may be zero if the upstream main street traffic does not cause a queue at downstream.

$$D_n = Ar(\Delta DEF)/h = \frac{|DE| \times |FG|}{2h} = \frac{\left\{d_n^g + \frac{(r_n - d_n^c)q_n^s}{q_{n+1}^d}\right\} \times (q_n^m t_n^q)}{2} \tag{3-12}$$

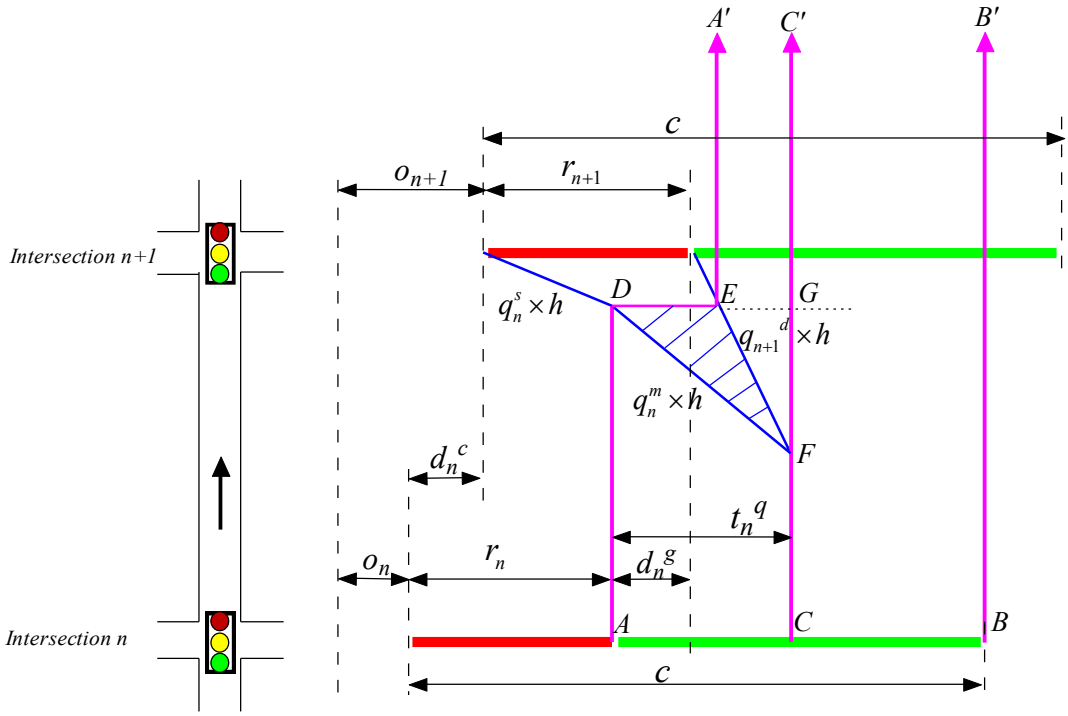


Figure 3-13 Queuing process between two intersections when $d_n^c \geq 0, d_n^g \geq 0$

b) $d_n^c < 0$

In this case, upstream red of the coordinated phase starts later than the downstream intersection. Vehicles discharged at the end of green of upstream intersection will cause a residual queue at downstream and the queuing curve is shown as HJ in Figure 3-14. After upstream coordinated phase turns red, vehicles from upstream side streets will keep increasing the downstream queue with an arrival rate of q_n^s. After upstream coordinated phase turns green, downstream intersection will have arrival vehicles from upstream coordinated phase with a rate of q_n^m (Veh/sec). If $d_n^g \geq 0$, as shown in Figure 3-14, upstream main street flow will cause a queue at downstream; if $d_n^g < 0$, upstream main street vehicles may or may not cause a queue at downstream depending on the arrival flow rate. According to flow conservation, equation (3-13) holds.

$$-d_n^c q_n^m + r_n q_n^s + t_n^q q_n^m = (t_n^q - d_n^g) q_{n+1}^d \tag{3-13}$$

t_n^q can be derived from equation (3-13) and $t_n^q = 0$ when upstream main street flow does not cause a queue at downstream. Therefore, t_n^q can be represented by

$$t_n^q = \max\left\{\frac{-d_n^c q_n^m + r_n q_n^s + d_n^g q_{n+1}^d}{q_{n+1}^d - q_n^m}, 0\right\} \tag{3-14}$$

The total delay of the main street traffic includes two parts, trapezium *HIKJ* and triangle *DEF*, as shown in the shadowed area in Figure 3-14. *HIKJ* represents the delay caused by residual vehicles discharged at the end of green of upstream coordinated phase and *DEF* is the delay caused by new arrival main street traffic. Note that the area of *DEF* could be zero if downstream queue is cleared before upstream main street vehicles arrive.

$$\begin{aligned} D_n &= Ar(HIKJ)/h + Ar(DEF)/h \\ &= \frac{[r_{n+1} + (r_n + d_n^g)] \times [-d_n^c q_n^m] + [-d_n^c q_n^m]^2 / q_{n+1}^d}{2} + \frac{\left\{d_n^g + \frac{r_n q_n^s + (-d_n^c q_n^m)}{q_{n+1}^d}\right\} \times (t_n^q q_n^m)}{2} \end{aligned} \tag{3-15}$$

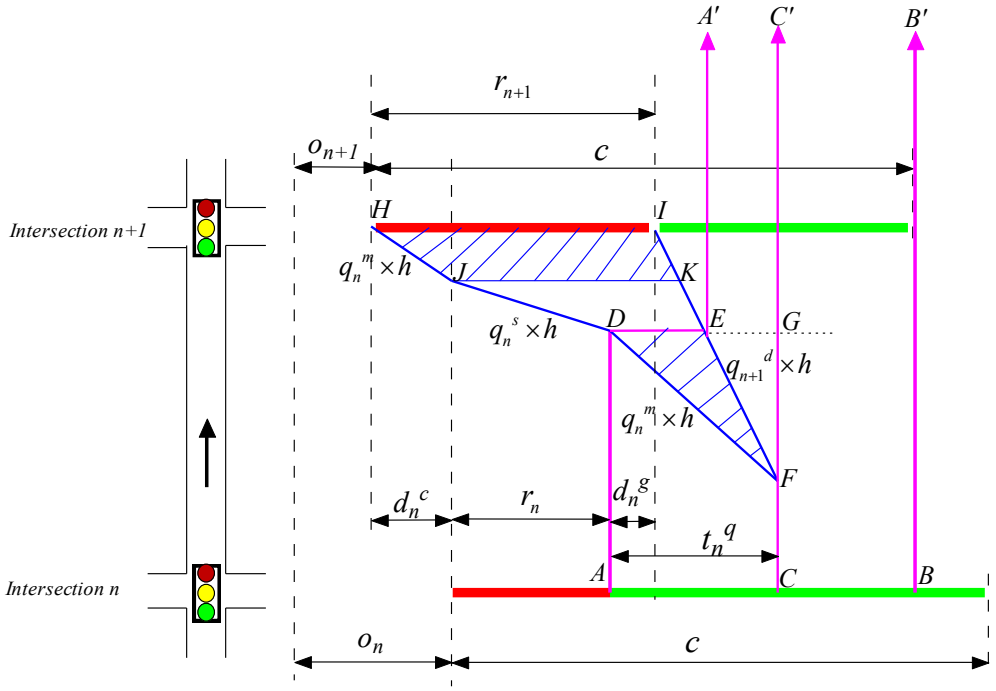


Figure 3-14 Queuing process between two intersections when $d_n^c < 0, d_n^g \geq 0$

(2) Extended Model

The basic delay calculation model discussed above assumes that vehicles are uniformly discharged during red and green periods of the coordinated phase (see Figure 3-13 and Figure 3-14). However, since the high-resolution traffic signal data timestamps each vehicle actuation, this assumption can be relaxed. We can actually get the discharging profile over the cycle for each approach at the boundary links and then derive the arrival profile for intermediate links.

We take the case when $d_n^c \geq 0$ as an example. At intersection *n*, the arrival flow rate per lane to downstream through links at any second τ_n of a cycle (denoted by $q_n(\tau_n)$, Veh./Sec./Ln.) can be generated by (3-16), where $w_{n,SB}^L(\tau_n)$, $w_{n,NB}^R(\tau_n)$ and $w_{n,EB}^T(\tau_n)$ represent the discharging flow rate (Veh./Sec.) of the corresponding movements of

intersection *n* at any second τ_n ($\tau_n \in \{1,2,...,c\}$). For boundary links, the discharging profile of each movement can be directly generated from the high-resolution traffic data. For example, if there are *y* cycles in a given scenario, there will be *y* discharging profiles for each movement of the boundary links. Each of these profiles will be used to generate the intersection delay. For intermediate links, the profile can be derived by (3-18), which will be discussed later. If we consider one cycle, the queuing process between two adjacent intersections is shown in Figure 3-15. At any second τ_n, the slope of the accumulated queuing curve is equal to $q_n(\tau_n)\times h$.

$$q_n(\tau_n) = \left(w^L_{n,SB}(\tau_n) + w^R_{n,NB}(\tau_n) + w^T_{n,EB}(\tau_n)\right)\varphi^T_{n+1,EB} / z_{n+1}, \tau_n \in \{1,2,...,c\} \quad (3\text{-}16)$$

According to flow conservation, equation (3-17) holds, from which t_n^q can be calculated.

$$\sum_{\tau_n=d_n^c}^{r_n+t_n^q} q_n(\tau_n) = (t_n^q - d_n^g)q_{n+1}^d \quad (3\text{-}17)$$

The discharging profile of the through movement at the downstream intersection *n+1* is then derived by (3-18). During the red light ($0 < \tau_{n+1} \le r_{n+1}$), the discharging flow rate is zero; during the queue discharging process ($r_{n+1} < \tau_{n+1} \le r_{n+1} + t_n^q - d_n^g$), the discharging flow rate is equal to the discharging flow rate per lane times number of lanes; after the queue is cleared, the discharging flow rate is equal to the corresponding arrival flow rate from upstream.

$$w^T_{n+1,EB}(\tau_{n+1}) = \begin{cases} 0 & 0 < \tau_{n+1} \le r_{n+1} \\ q_{n+1}^d z_{n+1} & r_{n+1} < \tau_{n+1} \le r_{n+1} + t_n^q - d_n^g \\ q_n\left((\tau_{n+1} + d_n^c) \bmod c\right) z_{n+1} & r_{n+1} + t_n^q - d_n^g < \tau_{n+1} \le c \end{cases} \quad (3\text{-}18)$$

The total delay of all the vehicles discharged from coordinated phase of intersection n can be calculated by (3-19). As shown in Figure 3-15, the area of DEF is equal to the area of DFG minus the area of the triangle EFG.

$$D_n = Ar(DEF)/h = \frac{Ar(DFG) - Ar(EFG)}{h}$$
$$= \sum_{u=r_n}^{r_n+t_n^q} \sum_{\tau_n=r_n}^{u} q_n(\tau_n) - \frac{q_{n+1}^d \times \left(t_n^q - d_n^g - \sum_{\tau_n=d_n^c}^{r_n} q_n(\tau_n)/q_{n+1}^d\right)^2}{2} \quad (3\text{-}19)$$

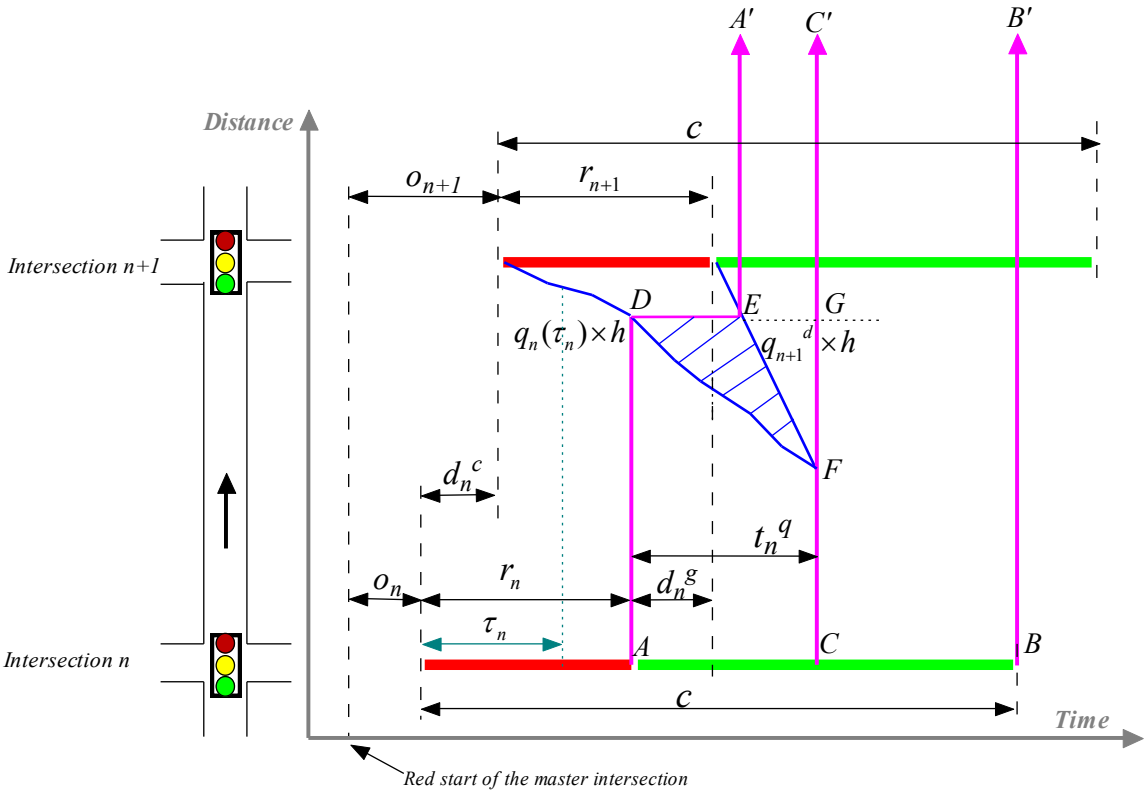


Figure 3-15 Queuing process between two intersections considering arrival profiles within a cycle when $d_n^c \geq 0, d_n^g \geq 0$

The calculation for the case when $d_n^c < 0$ can follow the same procedure. Based on the analysis above, the control delay for coordinated phases between each two adjacent intersections can be calculated accordingly given intersection signal timings and traffic scenario. Therefore, the scenario-based offset optimization model is complete by taking the delay calculation models into (3-5). For the sake of ease, we choose the basic delay calculation model in the following testing.

3.4 Solution Method and Optimization Results

Since the feasible region of intersection offsets, i.e. Ω, is a closed convex set (see Equation(3-7)) and the objective function is continuous and bounded within this region, an optimal solution to the problem must exist. Due to the complexity of the objective function, Genetic Algorithm (GA) is adopted to solve the problem. Many previous studies, such as Memon and Bullen (1996), Park et al. (2001), and Day and Bullock (2011), have also shown the effectiveness of GA when solving signal optimization problems. To implement a GA, a population of solutions (chromosomes) is first generated according to a creation rule, typically, in a random fashion. And then, the algorithm modifies the population of individual solutions repeatedly. At each step, depending on chromosome fitness, the algorithm selects chromosomes randomly from the current population to be parents and every two parent chromosomes will produce two new offspring chromosomes based on a cross-over rule. The process repeats itself until certain stopping criteria is met, at which time the best solution remains in the population become the final solution.

10 weekdays' (6/15/2009 ~ 6/26/2009) morning peak (7:00AM ~ 9:00 AM) data of 6 intersections along the TH55 were used to generate the optimal offsets. In the GA solution process, since we can take the first intersection as the master intersection with an offset of 0, each chromosome would have the length of 5, where each gene indicates the offset value of corresponding intersection. For the sake of simplicity, we take $\beta = 0$,

which means there is no trimming ($J_\beta = J$) and we consider all traffic scenarios. The weighting term α is set to 0.1, meaning that 10 percent of weight has been assigned to the performance of minor direction in our objective function. Using a GA toolbox in Matlab with following parameters shown in Table 3-1, a solution can be achieved after 40 generations in this case. The computing time is about 5 minutes using a desktop computer with Intel Xeon CPU W3540 (2.93 GHz) and 12 GB of RAM.

Table 3-1 Parameters for GA

Population Size	**Scaling Function**	**Elite Count**	**Crossover Fraction**	**Crossover Function**	**Mutation Function**	**Mutation Rate**
400	Rank	2	0.8	Two point	Uniform	0.1

Figure 3-16 shows the GA results at each generation, where the star "*" shows the mean function value among all individuals at corresponding generation and "." represents the function value of the best-fitted individual. Finally, the program generates the offset result {0, -21.4, -21, -20.9, -21.1, -21.2} for six intersections from Boone Ave. to TH100 respectively, comparing with the field offset setting {0, -24.9, -21.6, 4.6, -20.5, -11.9} under the transformed time coordinates. The final optimized offset values applied to controllers are {0, 8, 25, 47, 79, 114} and this offset setting generates a weighted total cycle delay of 304 (Veh.Sec), comparing with 638 (Veh.Sec) under the field offset setting.

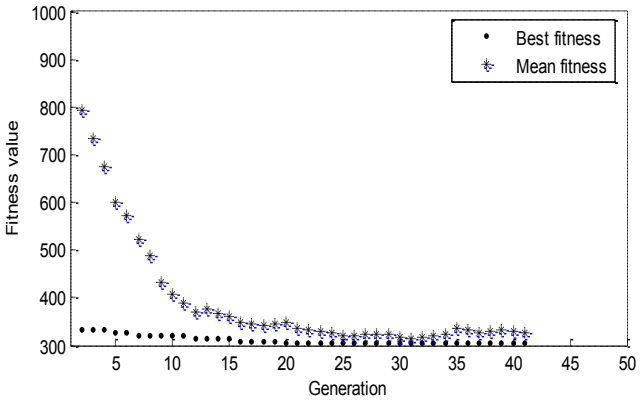


Figure 3-16 GA results for each generation

Figure 3-17 shows the coordination result for both field and optimized offset setting in the time-space diagram under the transformed time coordinates. The two figures on the left hand side show the existing field implementation for both phase 2 (eastbound) and phase 6 (westbound) and the two on the right show the optimized result. The arrows show the traffic direction for corresponding phases. Since original field offsets were already optimized by Synchro (the metropolitan district of Minnesota Department of Transportation retimes its traffic signals on all major arterials every 36 months or less), the generated offset values are very similar with the field ones. Major changes happen to Glenwood/TH55 intersection and TH100/TH55 intersection with a 25.5 and 9.3 seconds' reduction respectively. It is hard to directly tell the difference from the coordination figure, but as we will demonstrate in the next section, the optimized offset will further improve the performance of the coordinated phase (eastbound of TH55) without deteriorating the performance of the opposite direction (westbound of TH55).

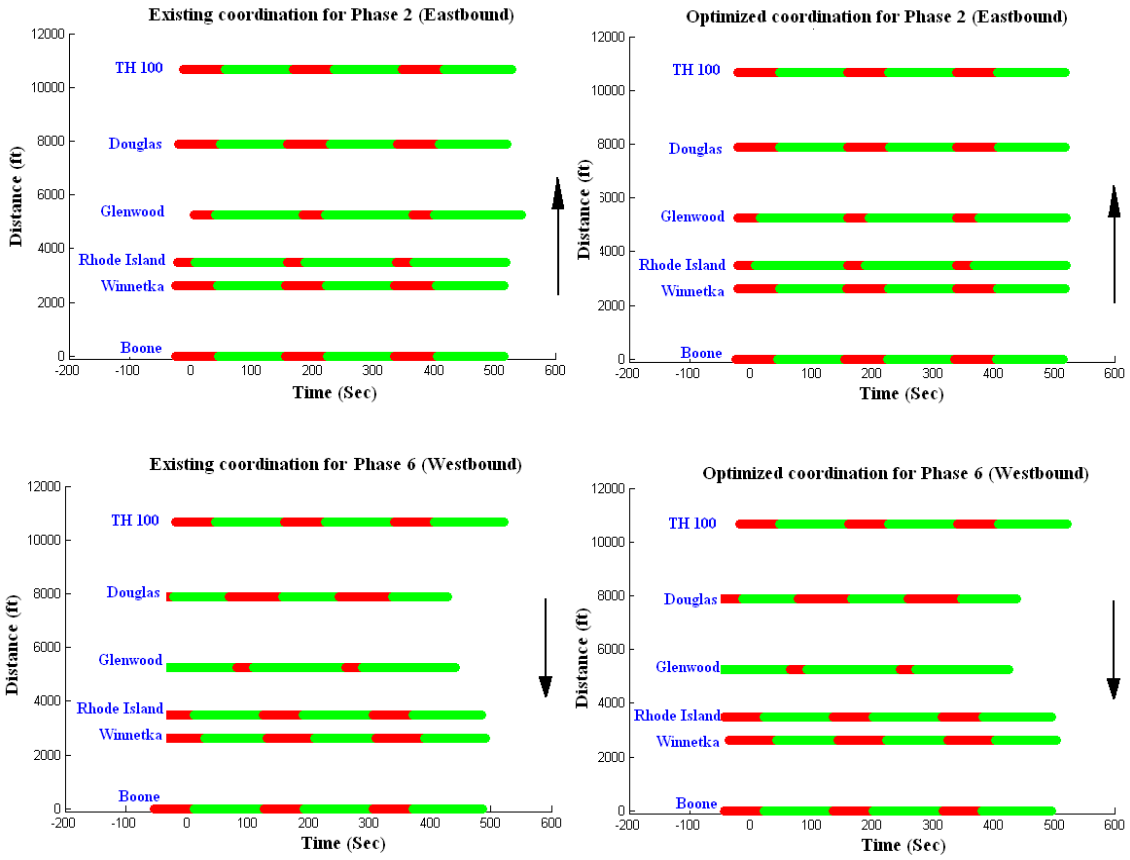


Figure 3-17 Coordination result for both field and optimized offsets

3.5 Field Evaluation

The optimized offset values were implemented in the 6 intersections on TH55 for morning peak hours (7:00AM~9:00AM) on 9/14/2009. In order to compare the performance measures before and after the field implementation, two days with similar volume inputs need to be chosen. After examining 10 days before and after, 9/3/2009 was selected as the baseline day with original offset setting to compare with 9/14/2009. Table 3-2 shows the comparison of Eastbound and Westbound (major direction) input volume during two hours' peak period and Table 3-3 shows the comparison for side street input volumes. As we can see, the two major directions' input volumes for the two days are

very similar and the day with optimized offset (9/14/2009) carries a little higher traffic. For side streets, these differences of input volumes are also within acceptable ranges.

Table 3-2 Eastbound and Westbound Input volume (7:00AM~9:00AM) comparison between 9/3/2009 and 9/14/2009

	9/3/2009 (Original offset)	**9/14/2009 (Optimized offset)**	**Change Percentage**
Eastbound input Volume	4615	4647	0.69%
Westbound input Volume	1817	1881	3.52%

Table 3-3 Side street Input volume (7:00AM~9:00AM) comparison between 9/3/2009 and 9/14/2009

		9/3/2009 (Original offset)	**9/14/2009 (Optimized offset)**	**Change Percentage**
Boone	Southbound	531	551	3.77%
	Northbound	445	452	1.57%
Winnetka	Southbound	748	675	-9.76%
	Northbound	285	269	-5.61%
Rhode	Southbound	456	470	3.07%
	Northbound	-*	-	-
Glenwood	Southbound	-	-	-
	Northbound	227	256	12.78%
Douglas	Southbound	943	934	-0.95%
	Northbound	213	207	-2.82%
TH 100	Southbound	586	622	6.14%
	Northbound	1008	923	-8.43%

*: T intersection, not applicable

Based on the high resolution data collected by the SMART-SIGNAL system, intersection queue length for each cycle at each of the intersections were estimated using the algorithm developed by Liu, et al. (2009). In order to compare the performance of the 6-intersection corridor before and after the implementation of the optimized offsets, arterial travel times for both directions were also calculated as part of the SMART-SIGNAL system based on the virtual probe algorithm developed by Liu and Ma (2009). Note the queue length estimation algorithm and the arterial travel time estimation algorithms implemented as part of the SMART-SIGNAL system has been tested and evaluated in the past and it is currently in operation on a number of intersections in Minnesota (for more information on the SMART-SIGNAL implementation, please see http://dotapp7.dot.state.mn.us/smartsignal). Since we only have 6 intersections, performance in between is what we are interested in. Therefore, travel time between stop line of Boone and stop line of TH100 is taken as an indicator of corridor performance.

Table 3-4 compares the calculated travel delays of both eastbound (from stop line of Boone to stop line of TH100) and westbound (from stop line of TH100 to stop line of Boone) based on different offset settings. As we can see, both eastbound and westbound travel delays are substantially reduced after the offset adjustment. On average, the eastbound travel delay with original offset (9/3/2009) is 11.98 seconds and it decreases to 10.14 seconds after optimization (9/14/2009), which is a 15.3% reduction. For westbound, average travel delay with original offset is 78.48 seconds and it decreases to 70.84 seconds after optimization, which indicates a 9.7% reduction. Considering that the original offset setting was already optimized, the improvement is significant. Figure 3-18 also compares the estimated eastbound queue length profiles during the one hour period at intersections 4 and 6 (where the major changes to offset happen) under different offset settings. One can find out that, in general, the optimized offset generates shorter queue lengths than the original offset. Therefore, the optimized offset outperforms the original offset in realistic field implementation.

Table 3-4 Eastbound and Westbound average delay comparison between 9/3/2009 and 9/14/2009

	9/3/2009 (Original)	9/14/2009 (Optimized)	Change Percentage
Eastbound Average Delay (Sec)	11.98	10.14	-15.3%
Westbound Average Delay (Sec)	78.48	70.84	-9.7%

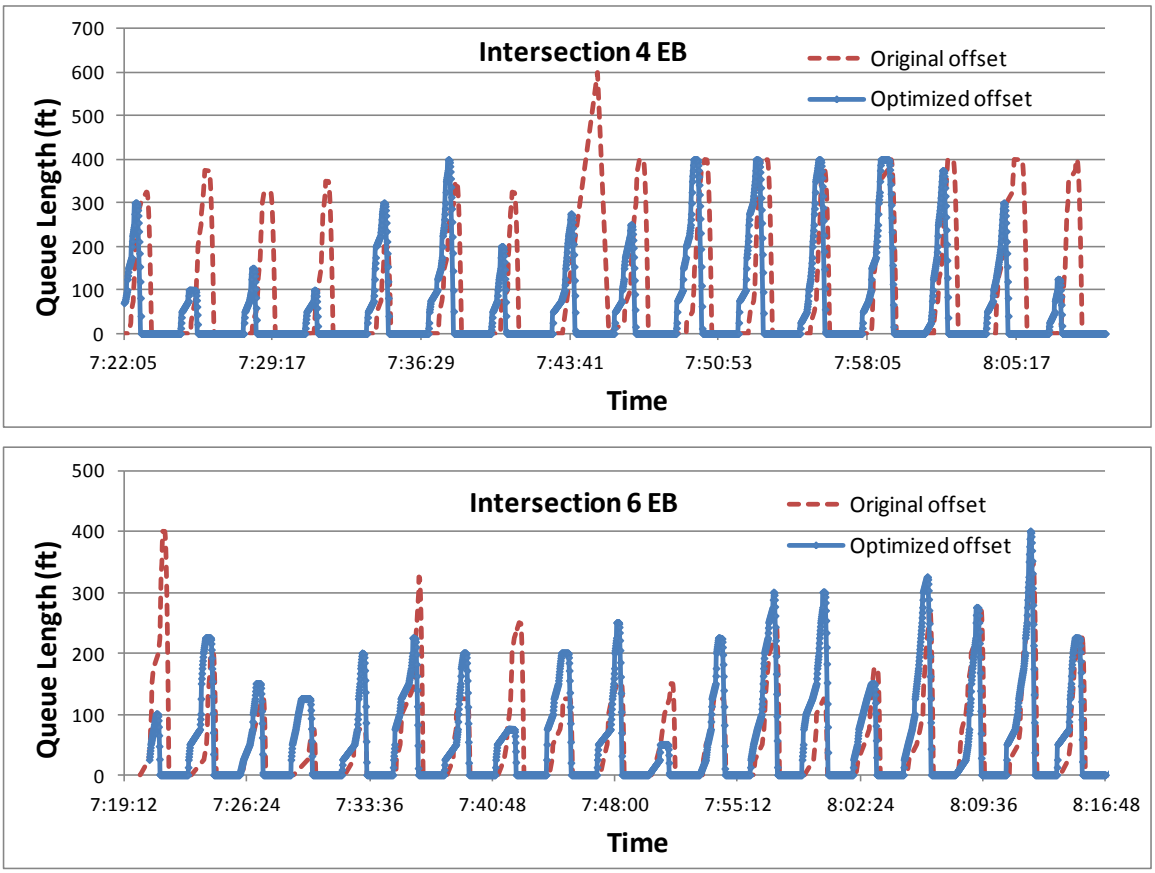


Figure 3-18 Estimated EB queue length based on different offset settings

3.6 Summary

In this chapter, an innovative approach for arterial offset optimization is developed based on the large amount of high-resolution field traffic data. The model solves the two well-known problems with actuated signals: (1) the early return to green problem for coordinated phases and (2) the uncertainty problem of queue sizes formed at intersections. To account for the two problems, we introduce the concepts of conditional distribution of the green start times and traffic demand scenarios, both of which can be obtained from the high-resolution traffic dataset. We explicitly consider the queues formed by side-street and main-street traffic under different situations. The objective of this model is to minimize total delay for one coordinated direction and at the same time take the performance of the other direction into consideration. Because of the complexity of the objective function, the problem is solved by the Genetic Algorithm. We implemented our optimized offset values to the field signal controllers and the test results indicate that the model successfully optimizes the offsets along the signalized arterials. It improves the performance of the corridor significantly.

4. Managing Oversaturated Signal Arterials: A Maximum Flow Based Approach

In Chapter 3, we discussed how to utilize the high-resolution traffic signal data to fine-tune the offsets and the field test has showed the effectiveness of the model. However, when traffic gets more and more congested, a new control strategy needs to be applied, since managing oversaturated arterials requires significantly different approaches for signal timing than those used for under-saturated conditions (Koonce et al., 2008). In this chapter, we will discuss our model to handle oversaturated arterials.

4.1 Background

Over the past fifty years, a fairly large body of literature has been devoted to this research area. For isolated intersections, optimal control policies were designed to balance the queuing delay on conflicting approaches by switching the green durations between minimum and maximum (e.g. Gazis & Potts, 1963 and Gazis, 1964; Michalopoulos & Stephanopoulos, 1977a, 1977b; Chang & Lin, 2000; etc.). To deal with oversaturated arterials, Pignataro et al. (1978) and Rathi (1988) proposed negative offset control strategy, which advances the downstream green in order to flush the residual queue. Lieberman et al. (2000) and Chang et al. (2010) developed traffic gating/metering control strategy, which aims to prevent downstream blockage by restraining upstream input. Liu and Chang (2011) presented a signal optimization model, which specifically considers the queue evolution by lane groups and can capture the queue spillback. At network level, Abu-Ledbdeh & Benekohal (1997), Park et al. (1999), and Lo et al. (2001) developed large-scale optimization models maximizing throughput with constraints on downstream storage capability and green time utilization.

However, in reality, none of these approaches have been widely accepted or implemented by practitioners because of the following reasons. Firstly, the basic assumption for most of the above models is that time-dependent traffic demand information is available. This is apparently problematic since most of the existing detection system fails to provide accurate arrival information, especially when traffic is congested. More importantly, the demand information, even it is available, cannot describe completely the level of oversaturation or the cause of oversaturation. Secondly, building large-scale optimization models is technically feasible but not very practical simply due to the complexity. It is almost impossible, especially for real-time traffic control, to solve large optimization programs with hundreds, or even thousands of decision variables and constraints; so most of these approaches cannot obtain an analytically optimal solution and have to rely on heuristic approaches. Finally, many practitioners do not feel comfortable to implement the "black-box" method provided by most of the optimization models.

We should note that adaptive signal control systems were also developed to perform real-time optimization of traffic signal splits, offsets and cycle lengths based on estimated or predicted traffic conditions, such as SCOOT (Hunt et al.,1982), SCATS (Sims, 1979), OPAC (Gartner, 1983), RHODES (Head et al., 1992) and ACS-Lite (Luyanda et al., 2003). It has been shown that these adaptive control systems can reduce delay in light to medium traffic conditions (Martin, 2007). When oversaturation happens, it is possible that these control systems may even worsen the situation due to the false estimation or prediction of traffic conditions (for example, arrival flow rate). Another impediment to adaptive traffic control systems is the cost associated with the deployment of additional detection systems that can supply the necessary traffic data for on-line decision-making.

To overcome the above difficulties, in this chapter, a simple but effective control model is developed to deal with oversaturation on signalized arterials. The control model is not based on time-dependent demand information, as most of the previous studies do. It is built upon the characterization of oversaturation discussed by Wu et al. (2010), in which,

oversaturation is quantified by the detrimental effects either in temporal dimension (indicated by a residual queue) or in spatial dimension (indicated by a spillover). The *oversaturation severity index* (*OSI*) indicates not only the level of congestion, but also the cause of oversaturation. With the information of *OSI*, we formulate a maximum flow based signal control model to manage oversaturation. Comparing with previous studies, this model has the following contributions:

(1) The proposed model can effectively and efficiently manage oversaturated routes without the requirement on time-dependent traffic demand as model inputs. It can be easily implemented at typical signalized intersections where the standard detection system is available.

(2) Instead of dealing with hundreds of decision variables and constraints, the control model can be solved by a simple Forward-Backward Procedure (FBP), which makes the practical use of the model much more promising.

In the following, Section 4.2 briefly reviews the definition of oversaturation severity index, which implies the possible mitigation strategies for oversaturation. Section 4.3 defines the problem we try to solve, followed by the maximum flow based formulation in Section 4.4. Section 4.5 introduces the FBP, which generates the optimal solution for the control model. The proposed model is evaluated through microscopic traffic simulation and the evaluation results are presented in Section 4.6. Some remarks are offered in the last section.

4.2 OSI-based Mitigation Strategies for Single Intersection

The model proposed in this chapter is built upon our recent developments on the diagnosis of oversaturation (Wu et al., 2010) using high-resolution traffic signal data (Liu & Ma, 2009; Liu et al., 2010). In Wu et al. (2010), an Oversaturation Severity Index (OSI) was proposed to quantify the severity level of oversaturation by measuring its detrimental effects. The detrimental effect is characterized by either a residual queue at the end of a cycle or a spillover from downstream traffic, both of which create “unusable”

green time. In the case of residual queue, the ‘‘unusable” green time is the equivalent green time to discharge the residual queue in the following cycle, but for spillover, the ‘‘unusable” green time is the time period during which a downstream link is blocked, therefore the discharge rate from the upstream intersection is zero. OSI is defined as the ratio between unusable green time and total available green time in a cycle, which is a non-negative percentage value between 0 and 100, with 0 indicating no detrimental effect for signal operation, and 100 being the worst that all available green time becomes unusable. OSI is further differentiated into *TOSI* (Temporal Oversaturation Severity Index, caused by the residual queue that creates the detrimental effect in temporal dimension) and *SOSI* (Spatial Oversaturation Severity Index, caused by the spillover that creates the detrimental effect in spatial dimension). Specifically, *TOSI* and *SOSI* can be calculated by Equation (4-1). More detailed derivation and explanation of the two indices can be found in Wu et al. (2010).

$$\begin{cases} TOSI = \dfrac{\text{Green time to discharge residual queue}}{\text{Total available green time}} \\ SOSI = \dfrac{\text{Unusable green time due to spillover}}{\text{Total available green time}} \end{cases} \tag{4-1}$$

With *TOSI* and *SOSI*, not only can the severity level of oversaturation be quantified, but also the causes of arterial traffic congestion can be identified. Positive *TOSI* indicates that the available green time is insufficient for queue discharge and a residual queue is formed at the end of a cycle; and positive *SOSI* indicates that the queue length at the downstream link has reached the upstream intersection and blocked the discharging traffic. Based on measured *TOSI* and *SOSI* values, three basic mitigation strategies are designed for different oversaturation scenarios between two intersections.

1) Green Extension for Scenario 1: TOSI > 0 & SOSI = 0.

Since a positive *TOSI* value indicates a residual queue at the end of a cycle and zero *SOSI* value indicates that there is still spare capacity to store vehicles in the downstream link,

the strategy to deal with this situation is to extend the green time for the oversaturated phase.

Figure 4-1 (a) illustrates this case by presenting the shockwave profiles for two intersections. After extending the green in

Figure 4-1 (b), the residual queue disappears and *TOSI* becomes zero. The green extension can be calculated as the following (Equation(4-2)).

$$\Delta g_{n,i} = TOSI_{n,i} \times g_{n,i} \qquad (4\text{-}2)$$

where $\Delta g_{n,i}$ is the adjustment to the green time at intersection *n* for phase *i*; $TOSI_{n,i}$ is the *TOSI* value at intersection *n* for phase *i* ; and $g_{n,i}$ is the green time at intersection *n* for phase *i*. Note that positive $\Delta g_{n,i}$ means green extension; and a negative value means green reduction. By extending green, the start time of the following red signal will be postponed.

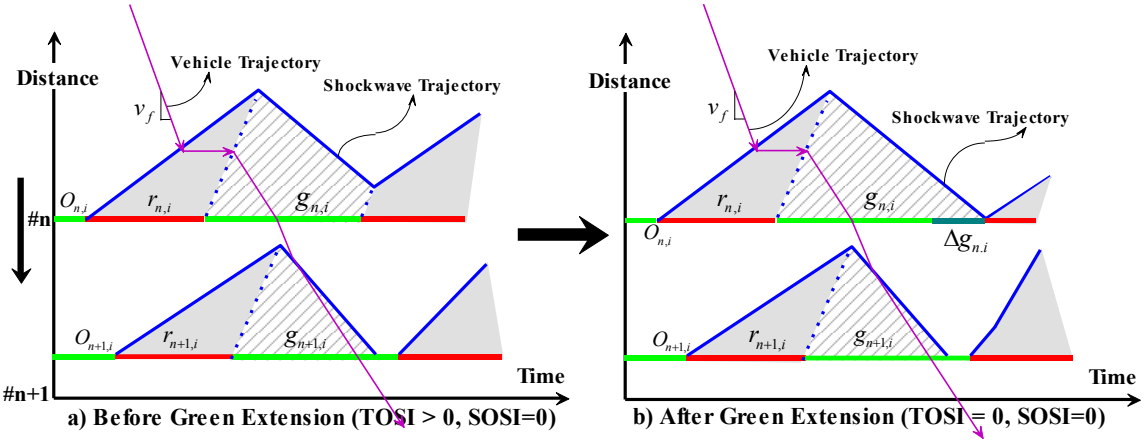


Figure 4-1 Green extension for Scenario 1

2) Red Extension for Scenario 2: TOSI = 0 & SOSI > 0.

If *SOSI* is larger than zero, it indicates that the downstream queue spills back to the upstream intersection and results in unusable green time as shown in Figure 4-2(a). But since *TOSI* is zero, all queued vehicles can be discharged even with reduced green time. One way to remove downstream spillover is to gate the upstream flow by extending the red time. The red extension can be calculated as the following (Equation(4-3)).

$$\Delta r_{n,i} = SOSI_{n,i} \times g_{n,i} \qquad (4\text{-}3)$$

Where $\Delta r_{n,i}$ is the adjustment to the red time at intersection *n* for phase *i*; and $SOSI_{n,i}$ is the *SOSI* value at intersection *n* for phase *i*. The positive $\Delta r_{n,i}$ means red extension and a negative value means red reduction. Note that by extending the red time, the start of the following green will be postponed.

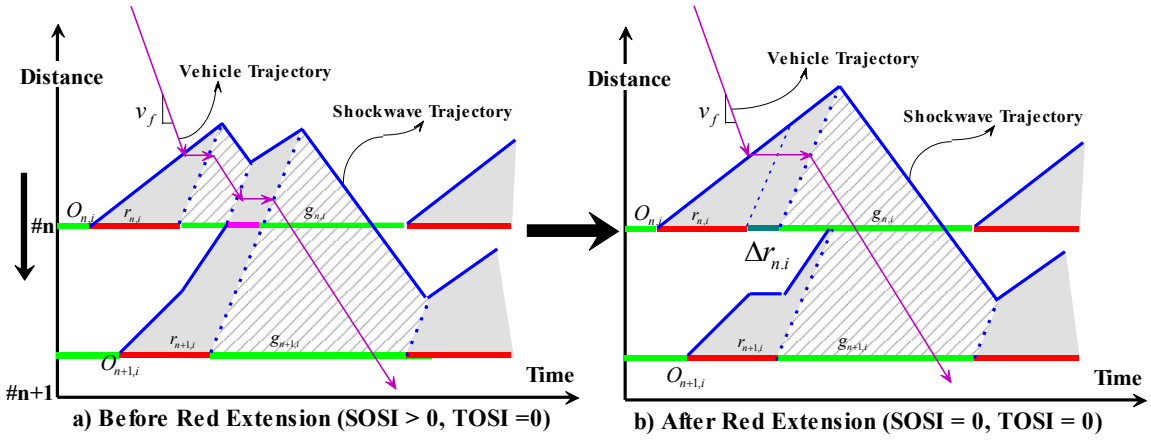


Figure 4-2 Red extension for Scenario 2

3) Downstream Red Reduction for Scenario 3: TOSI > 0 & SOSI > 0.

A more serious situation exists when both *TOSI* and *SOSI* are larger than zero, as shown in Figure 4-3(a). In this case, at the upstream intersection a portion of the green time is unused because of the downstream spillover. At the same time, the usable green time at the upstream intersection is not sufficient to discharge queued vehicles, i.e., a residual

queue exists. One way to deal with this scenario is to increase downstream capacity by reducing the red time at the downstream intersection. As shown in Figure 4-3, by reducing the downstream red, positive *TOSI* and *SOSI* values for the upstream intersection can be reduced. Once the downstream spillover is removed or reduced, the unusable green time at the upstream intersection may become available and can be used to discharge the residual queue. If *TOSI* < *SOSI*, the residual queue can be cleared by using this strategy. The reduction of downstream red can be calculated as the following (Equation (4-4)).

$$\Delta r_{n+1,i} = -SOSI_{n,i} \times g_{n,i} \qquad (4\text{-}4)$$

It should be noted, as an alternative to reduce the downstream red light time, we can also deal with situation by combining the methods for scenario 1 and 2 together, i.e., extending both the red and green times.

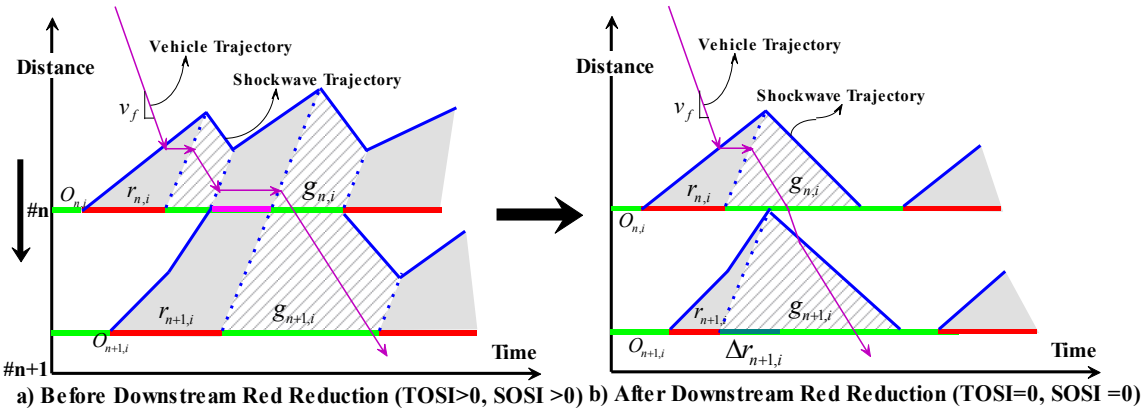


Figure 4-3 Red reduction (at downstream intersection) for Scenario 3

Among the three strategies, extending green (strategy 1) is to increase the discharge capacity for the oversaturated phase; extending red (strategy 2) is to gate traffic arrivals at the upstream intersection; and reducing downstream red (strategy 3) is to remove the downstream bottleneck by discharging the queue earlier at downstream intersection. By considering maximum/minimum green and storage space limitations on side streets, the

strategies introduced above may be directly applied for an isolated intersection or two intersections in tandem. However, if oversaturation occurs on multiple intersections along a signalized arterial, a systematic strategy will be needed and this will be introduced next.

4.3 Maximum Flow Problem Formulation for Arterial Oversaturation

When oversaturation occurs on multiple intersections along a signalized arterial, the strategies introduced above cannot be applied directly because: 1) the increase of green time of an upstream approach may create oversaturation on the downstream link and, 2) capacity constraints at a downstream phase may limit the possible signal timing adjustments for the upstream phase. Therefore a systematic approach will be needed. In this section, we formulate the arterial oversaturation problem as a maximum flow problem.

In our formulation, we assume that the oversaturation severity indices for each of the intersection along the arterial are available (see Figure 4-4). In other words, *TOSI* and *SOSI* values for each phase *i* of every intersection *n* are measured at the end of every cycle (see Wu et al., (2010) for details on calculating *TOSI* and *SOSI*) , and at least some of these *TOSI* and *SOSI* values are positive.

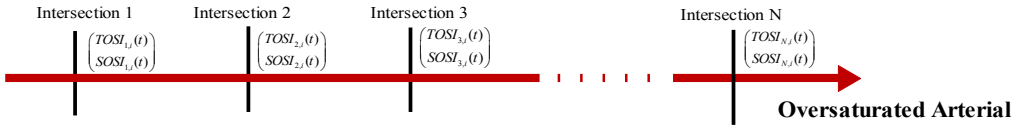


Figure 4-4 Intersections along an oversaturated arterial

Now the problem becomes that, at the end of control period $t-1$, given the oversaturation severity indices measured in this period, i.e., $TOSI_{n,i}(t-1)$ and

$SOSI_{n,i}(t-1)$, determines the signal timing changes (cycle lengths, green splits, and offsets) for each intersection at control period t to maximize the discharging capacity. We assume that each control period contains at least one cycle so that *TOSI* and *SOSI* can be measured. For the ease of real-world implementation of the proposed model, we also assume that cycle lengths among all intersections along the oversaturated arterial remain unchanged.

4.3.1 Control Variables

In the maximum flow model, two sets of control variables $\Delta r_{n,i}(t)$ and $\Delta g_{n,i}(t)$, namely red time changes and green time changes for phase i at intersection n, are introduced for each oversaturated phase. The two control variables have direct association with specific oversaturation mitigation strategies. Whether to change red or green is determined by the causes of the oversaturation. Changing red times (i.e. $\Delta r_{n,i}$) aims to eliminate spillover; and changing green times (i.e. $\Delta g_{n,i}$) aims to clear residual queues. A positive red time change (**red extension**) means that extra red time is added. Since the cycle length is kept unchanged, the green start would be postponed with the red extension (see Figure 4-5a) and the total green time is reduced. A negative red time change (**red reduction**) means a portion of red time is cut from the end of red, therefore, green start will be advanced (see Figure 4-5b) and the total green time is increased. Similarly, a positive green time change (**green extension**) indicates that additional green time is added to the original end of the green time (see Figure 4-5c), and a negative green time change (**green reduction**) represents that some green time is cut from the end of green (see Figure 4-5d). Depending on the offset reference point used for the intersection (start of yellow, start of green, barrier crossing, etc.), each case of adjusting green or red may require a corresponding change to the offset and green split values.

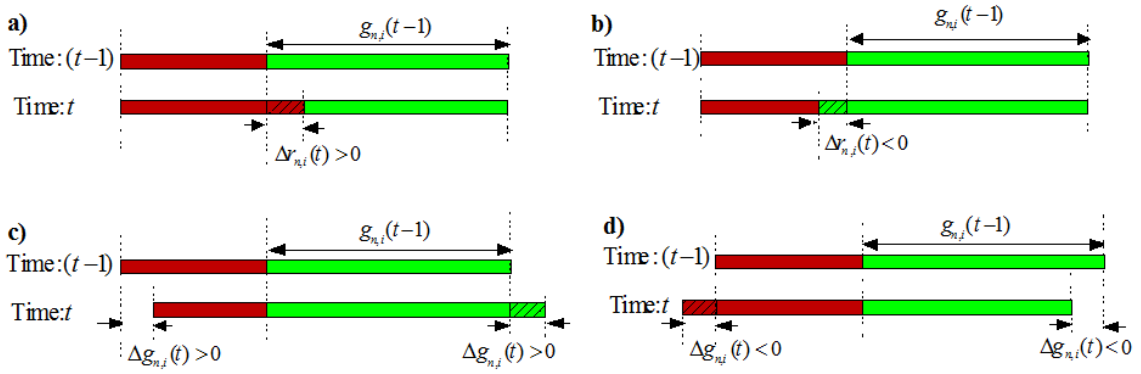


Figure 4-5 Red Time Changes & Green Time Changes

The values $\Delta r_{n,i}(t)$ and $\Delta g_{n,i}(t)$ can be easily transformed into the values of new offset and green duration, which can easily be modified in the signal timing plan. If we assume that the oversaturated phase is the coordinated direction and the red start time of the coordinated phase is the offset reference point, Equation(4-5) can be used to calculate the new offset and green duration after adjustment, where $o_n(t)$ is the offset value at time t and c_n is the cycle length for intersection n. Figure 4-6 presents an example of signal timing changes at one intersection with $\Delta r_{n,i}(t) < 0$ and $\Delta g_{n,i}(t) > 0$. Note that, if there is no change on the offset, the designed signal timing can be achieved in the immediately next cycle; however, if there is a change on the offset, a transition period may be needed in order to get to the new timing.

$$\begin{cases} o_n(t) = o_n(t-1) + \Delta g_{n,i}(t) \\ g_{n,i}(t) = g_{n,i}(t-1) - \Delta r_{n,i}(t) + \Delta g_{n,i}(t) \\ r_{n,i}(t) = c_n - g_{n,i}(t) \end{cases} \quad (4\text{-}5)$$

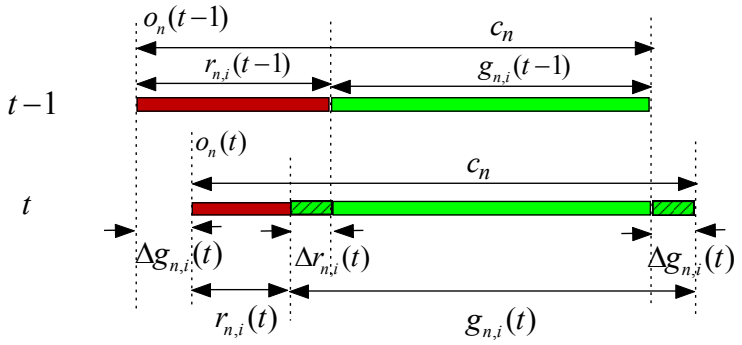


Figure 4-6 Signal timing changes ($\Delta r_{n,i}(t) < 0,\ \ \Delta g_{n,i}(t) > 0$)

4.3.2 Constraint Analysis

In the following, for the sake of simplicity, we use $S_{n,i}(t)$ to represent the unusable green time caused by spillover (i.e., *SOSI* > 0) at intersection *n* for phase *i* during time period of *t*, and use $T_{n,i}(t)$ to represent the unusable green time caused by residual queue (i.e., *TOSI* > 0).

$$\begin{cases} S_{n,i}(t) = SOSI_{n,i}(t) \times g_{n,i}(t) \\ T_{n,i}(t) = TOSI_{n,i}(t) \times g_{n,i}(t) \end{cases} \tag{4-6}$$

1) Spillover Elimination

The proposed control model aims to eliminate spillover between intersections. As presented previously, in order to eliminate the spillover at intersection *n*, one can either extend the red time at the current intersection *n* (i.e., apply gating at the upstream intersection, see Figure 4-2), or reduce the red time at the downstream intersection *n+1* (i.e., discharge the downstream queue earlier, see Figure 4-3), or a combination of the two strategies. As described in Figure 4-7, extending the red at intersection *n* by $\Delta r_{n,i}(t)$ ($\Delta r_{n,i}(t) > 0$) will make the unusable green time caused by spillover shorter by

$\Delta r_{n,i}(t)$; On the other hand, reducing the red at intersection $n+1$ by $\left|\Delta r_{n+1,i}(t)\right|$ ($\Delta r_{n+1,i}(t) < 0$) will make the unusable green caused by spillover at intersection n shorter by $\left|\Delta r_{n+1,i}(t)\right|$. Therefore, in order to eliminate spillover at intersection n, the difference of red time changes between intersection n and intersection $n+1$ should be equal to the unusable green time caused by spillover at intersection n, i.e., $S_{n,i}(t-1)$, see Equation (4-7).

$$\Delta r_{n,i}(t) - \Delta r_{n+1,i}(t) = S_{n,i}(t-1), \forall n \in \{1,\ldots,N-1\} \tag{4-7}$$

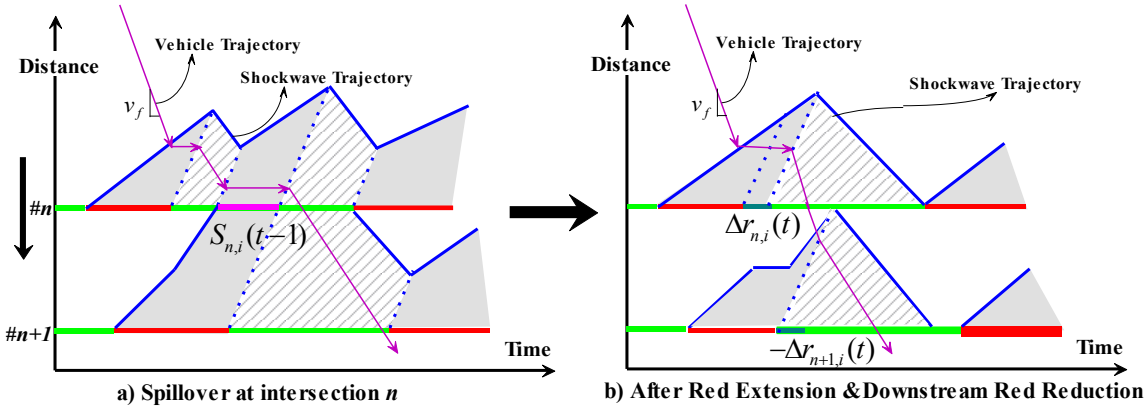


Figure 4-7 An example of applying strategies 2 & 3 to eliminate spillover

2) Residual Queue Elimination

If Equation (4-7) is satisfied, the spillovers are supposed to be eliminated during control period t. Then the green time change $\Delta g_{n,i}(t)$ for each intersection is used to eliminate residual queue. If the red time and green time changes at intersection n are $\Delta r_{n,i}(t)$ and $\Delta g_{n,i}(t)$ respectively, the total green time at intersection n for control period t would be $[-\Delta r_{n,i}(t) + \Delta g_{n,i}(t) + g_{n,i}(t-1)]$. If Intersection $n+1$ has residual queue in control period

$t-1$ and the corresponding unusable green time is $T_{n+1,i}(t-1)$, in order to eliminate residual queue of Intersection $n+1$ at control period t, the difference of total green time between Intersection *n+1* and its upstream intersection *n* should be equal to $T_{n+1,i}(t-1)$, i.e., Eq. (4-8) should hold.

$$\begin{aligned}&\left[\Delta g_{n+1,i}(t)-\Delta r_{n+1,i}(t)+g_{n+1,i}(t-1)\right]-\left[\Delta g_{n,i}(t)-\Delta r_{n,i}(t)+g_{n,i}(t-1)\right]\\&=T_{n+1,i}(t-1), n\in\{1,...,N-1\}\end{aligned} \tag{4-8}$$

Substitute (4-7) into (4-8),

$$\begin{aligned}&\Delta g_{n+1,i}(t)-\Delta g_{n,i}(t)\\&=T_{n+1,i}(t-1)-S_{n,i}(t-1)-\left[g_{n+1,i}(t-1)-g_{n,i}(t-1)\right], n\in\{1,...,N-1\}\end{aligned} \tag{4-9}$$

3) Available Green Constraints

For each intersection along the oversaturated route, the green time increase at control period t, i.e., $\Delta g_{n,i}(t)-\Delta r_{n,i}(t)$ is constrained by the available green time $g_{n,i}^{a}(t)$ for intersection *n* and phase *i*, see Eq.(4-10).

$$\Delta g_{n,i}(t)-\Delta r_{n,i}(t)\le g_{n,i}^{a}(t), \forall n\in\{1,...,N\} \tag{4-10}$$

If $Z_{n,i}$ is the set of conflicting phases to phase *i* at intersection *n*, the available green time $g_{n,i}^{a}(t)$ can be computed by considering the maximum queue size for each of these conflicting phases in the immediate past control interval $t-1$, see Eq.(4-11). Here c_n is the cycle length for intersection *n*, $Q_{n,p}^{\max}(t-1)$ is the maximum queue size per lane for phase *p* at intersection *n* at control interval $t-1$ and $s_{n,p}$ is the saturation flow rate per

lane for phase *p* of intersection *n*. $Q_{n,p}^{\max}(t-1)/s_{n,p}$ calculates how much green time is needed to discharge the queue of $Q_{n,p}^{\max}(t-1)$. α^{Q} is a weighting term, which represents users' perspective on the importance of queues on conflicting phases when calculating the available green for oversaturated phase i. When the maximum queue length for phase *p, i.e.* $Q_{n,p}^{\max}(t-1)l$ (where l is the space headway under jammed traffic condition), is shorter than the corresponding link length $L_{n,p}$, we may account for only a portion ($0 \le \alpha^{Q} < 1$) of these queues because we want to maximize the discharging capacity for the oversaturated route to reduce congestion; however, if the maximum queue length for phase *p* is already longer than the link length, all the queues need to be considered ($\alpha^{Q} = 1$), otherwise these queues will block further upstream intersections. One should note that, the smaller the α^{Q} is, the more expected extra capacity will be assigned to the oversaturated route, the faster the queues on conflicting phases will grow and the more delay will be introduced to conflicting phases.

$$g_{n,i}^{a}(t) = c_n - \sum_{p \in Z_{n,i}} \alpha^{Q} \left[Q_{n,p}^{\max}(t-1)/s_{n,p} \right] - g_{n,i}(t-1) \qquad \text{(4-11)}$$

where $0 \le \alpha^{Q} \le 1$.

4.3.3 Maximum Flow Based Control Model

1) One oversaturated route

The objective of the control model is to maximize the discharging capacity along the oversaturated route. At each control period *t*, it is equivalent to maximizing the total green time at the first intersection of the route, i.e., $\left(\Delta g_{1,i}(t) - \Delta r_{1,i}(t) + g_{1,i}(t-1)\right)$. Since $g_{1,i}(t-1)$ is the green time during control period $t-1$, at the start of control period *t*, maximizing $\left(\Delta g_{1,i}(t) - \Delta r_{1,i}(t) + g_{1,i}(t-1)\right)$ is equivalent to maximizing $\left(\Delta g_{1,i}(t) - \Delta r_{1,i}(t)\right)$.

Therefore, the complete control model can be expressed in (4-12). The first and second constraints ensure the elimination of spillover and residual queues between intersections and the third constraint considers the available green time. Note that, if the green duration for the oversaturated route is shortened at some intersection due to red extension or green reduction, the non-oversaturated directions will receive more green time and that may cause more vehicles from the non-oversaturated directions to be added to the oversaturated route because of turning. However, in most cases, the traffic volume discharged from the non-oversaturated directions is much smaller than the one from the oversaturated direction, so the strategy will not worsen the traffic condition. Further, if this strategy makes the previously non-oversaturated route become oversaturated, the traffic condition will then be considered as the case with two intersecting oversaturated routes, as will be discussed in the next section.

$$
\begin{aligned}
&\max \Delta g_{1,i}(t) - \Delta r_{1,i}(t) \\
&s.t. \\
&\Delta r_{n,i}(t) - \Delta r_{n+1,i}(t) = S_{n,i}(t-1), && n \in \{1,...,N-1\} \\
&\Delta g_{n+1,i}(t) - \Delta g_{n,i}(t) && \\
&\quad = T_{n+1,i}(t-1) - S_{n,i}(t-1) - \left[g_{n+1,i}(t-1) - g_{n,i}(t-1) \right], && n \in \{1,...,N-1\} \\
&\Delta g_{n,i}(t) - \Delta r_{n,i}(t) \le c_n - \sum_{p \in Z_{n,i}} \alpha^{Q} \left[Q_{n,p}^{\max}(t-1) / s_{n,p} \right] - g_{n,i}(t-1), && n \in \{1,...,N\}
\end{aligned}
\tag{4-12}
$$

If we treat the signal timing changes of each oversaturated phase along the route as flows in a network, the above linear program can be seen as a multi-commodity maximum flow problem. The corresponding network $G(M, A)$ is shown in Figure 4-8 with node set *M* and arc set *A*. Each intersection $n \in \{1,...,N\}$ along the oversaturated route is corresponding to a node *n* in the network. The node set *M* also includes two dummy nodes: the source node S_i and the sink node D_i. The arc coming out of node $n \in \{1,...,N\}$ has a capacity constraint $g_{n,i}^{a}(t)$, which limits the total flow it can carry.

There are two kinds of flows or commodities in the network, namely red reduction $-\Delta r_{n,i}(t)$ and green extension $\Delta g_{n,i}(t)$. In Figure 4-8, we place the flow of red reduction on the upper side of the corresponding arc and the green extension on the lower side. The total flow of the two commodities on each arc is constrained by the arc capacity, i.e., $\Delta g_{n,i}(t) - \Delta r_{n,i}(t) \le g^a_{n,i}(t)$, $\forall n \in \{1,...,N\}$. At each node $n \in \{2,3,...,N\}$, there is an input flow $f^r_{n,i}(t)$ for red reduction and another input flow $f^g_{n,i}(t)$ for green extension. The definition of $f^r_{n,i}(t)$ and $f^g_{n,i}(t)$ is shown in Equation(4-13). From a network flow point of view, at each node, the two input flows are external demands for the two commodities, respectively.

$$\begin{cases} f^r_{n+1,i}(t) = S_{n,i}(t-1) \\ f^g_{n+1,i}(t) = T_{n+1,i}(t-1) - S_{n,i}(t-1) - \left[g_{n+1,i}(t-1) - g_{n,i}(t-1) \right] \end{cases}, \quad n \in \{1,...,N-1\} \qquad (4\text{-}13)$$

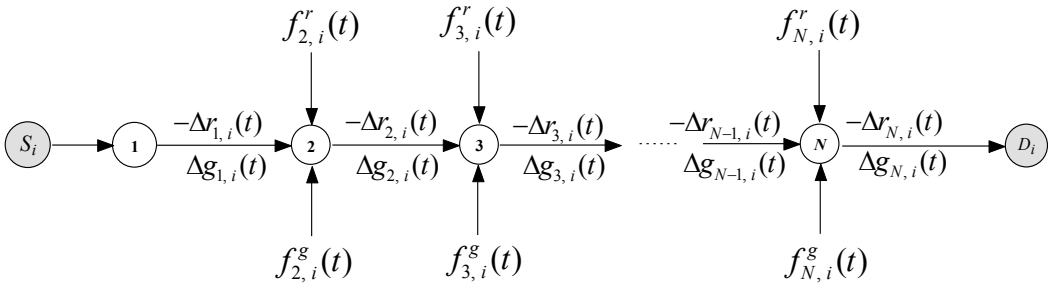


Figure 4-8 Maximum flow network for one oversaturated route

2) Two intersecting oversaturated routes

If there are two intersecting oversaturated routes, similar approach can be applied to construct a maximum flow problem. As shown in Figure 4-9, Route *1* includes intersection *1*, intersection *2*, …, intersection *N*, and follows the direction of phase *i*; Route *2* includes intersection *1'*,intersection *2'*, …, intersection *N'*, and follows the direction of phase *j*. Two oversaturated routes intersect with each other at the critical intersection *I*.

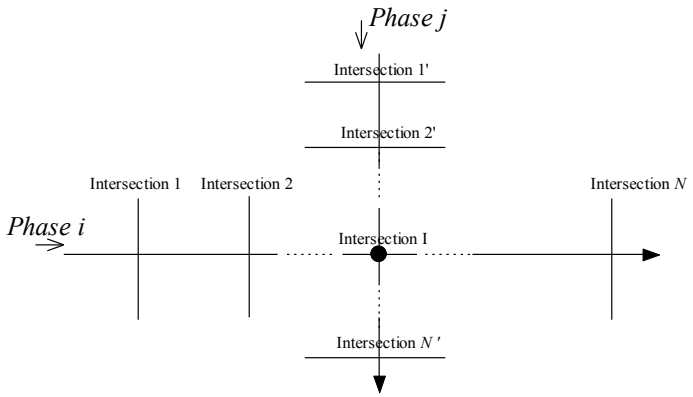


Figure 4-9 Two intersecting oversaturated routes

For two intersecting oversaturated routes, the objective of the control model is to maximize the total flows for both routes while satisfying spillover elimination, residual queue elimination and flow-capacity constraints. At intersection *I*, the available green for both phase *i* and *j* ($g^{a}_{I,\,i\&j}(t)$) can be calculated by Eq.(4-14), where $Z_{I,\,i\&j}$ is the set of conflict phases to phase *i* and *j* at intersection *I*.

$$g^{a}_{I,\,i\&j}(t) = c_I - \sum_{p \in Z_{I,\,i\&j}} \alpha^{Q} \left[Q^{\max}_{I,p}(t-1) / s_{I,p} \right] - g_{I,i}(t-1) - g_{I,j}(t-1) \qquad (4\text{-}14)$$

The control model can be formulated in the form of (4-15). Most constraints have similar meanings as what we described for Eq.(4-12), except for the last one, which is the available green constraint at intersection *I*.

$$
\begin{aligned}
&\max\left(\Delta g_{1,i}(t)-\Delta r_{1,i}(t)\right)+\left(\Delta g_{1',j}(t)-\Delta r_{1',j}(t)\right)\\
&s.t.\\
&\Delta r_{n,k}(t)-\Delta r_{n+1,k}(t)=S_{n,k}(t-1),\\
&\quad n\in\{1,\ldots,N-1\}\&k=i \text{ or } n\in\{1',\ldots,N'-1\}\&k=j\\
&\Delta g_{n+1,k}(t)-\Delta g_{n,k}(t)=T_{n+1,k}(t-1)-S_{n,k}(t-1)-\left[g_{n+1,k}(t-1)-g_{n,k}(t-1)\right]\\
&\quad n\in\{1,\ldots,N-1\}\&k=i \text{ or } n\in\{1',\ldots,N'-1\}\&k=j\\
&\Delta g_{n,k}(t)-\Delta r_{n,k}(t)\le c_n-\sum_{p\in Z_{n,k}}\alpha^{Q}\left[Q_{n,p}^{\max}(t-1)/s_{n,p}\right]-g_{n,k}(t-1),\\
&\quad n\in\{1,\ldots,I-1\}\cup\quad,N\}\&k=i\\
&\quad \text{or } n\in\{1',\ldots,(I-1)'\}\cup\quad\ldots,N'\}\&k=j\\
&-\Delta r_{I,i}(t)-\Delta r_{I,j}(t)+\Delta g_{I,i}(t)+\Delta g_{I,j}(t)\\
&\quad\le c_I-\sum_{p\in Z_{I,i\&j}}\alpha^{Q}\left[Q_{I,p}^{\max}(t-1)/s_{I,p}\right]-g_{I,i}(t-1)-g_{I,j}(t-1)
\end{aligned}
\tag{4-15}
$$

The corresponding maximum flow network is presented in Figure 4-10. There are two source nodes S_i and S_j, and two sink nodes D_i and D_j, for two oversaturated routes respectively. For each direction, it has similar network structures as Figure 4-8, except the critical node *I*, which is impacted by both routes. At this special node, flows from both directions need to be considered. To deal with this, we re-design the maximum flow network, in which, original node *I* is represented by two nodes *I* and $\tilde{I}$ (see Figure 4-10), and a new arc ($I,\tilde{I}$) is added. The capacity of arc ($I,\tilde{I}$) is constrained by the available green time for intersection (node) *I*, i.e., $g^{a}_{I,i\&j}(t)$. At node *I*, there are two pairs of incoming flows $-\Delta r_{I-1,i}(t)\ \&\ \Delta g_{I-1,i}(t)$ and $-\Delta r_{(I-1)',j}(t)\ \&\ \Delta g_{(I-1)',j}(t)$ from node *I-1* and *(I-1)'* respectively; there are external flows $f^{r}_{I,i}(t)$ and $f^{r}_{I,j}(t)$ for red reductions of two directions and external flows $f^{g}_{I,i}(t)$ and $f^{g}_{I,j}(t)$ for green extensions of two directions. The total flow coming out of node *I* is the summation of flows for both

directions, $-\Delta r_{I,i}(t)-\Delta r_{I,j}(t)+\Delta g_{I,i}(t)+\Delta g_{I,j}(t)$, which is constrained by the capacity of arc $(I,\tilde{I})$, i.e., $g^{a}_{I,i\&j}(t)$.

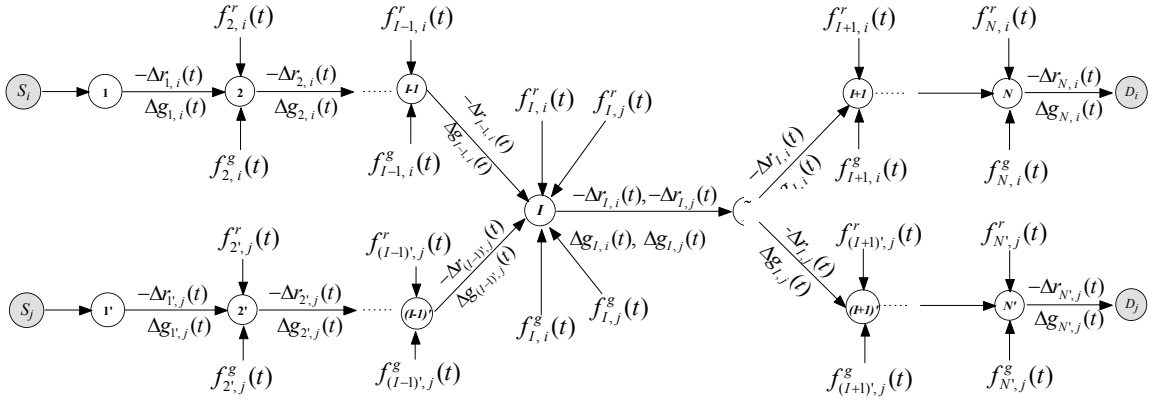


Figure 4-10 Maximum flow network for two intersecting oversaturated routes

4.3.4 Model Discussion

1) Solution existence

It is not difficult to find out that (4-12) always has a feasible solution. However, the formulation actually ignores one constraint, where the green duration of each intersection cannot be negative. In other words, $\Delta g_{n,i}(t)-\Delta r_{n,i}(t)+g_{n,i}(t-1)\geq 0$ should hold at each intersection. But this constraint may be violated under some extreme conditions when the spillover time or residual queue length is too large. To overcome this problem, slack variables $\varepsilon^{S}_{n,i}(t)$ and $\varepsilon^{T}_{n,i}(t)$ are introduced. They represent the unsolved portion of $S_{n,i}(t-1)$ and $T_{n+1,i}(t-1)$ respectively in control period *t*. Therefore, the summation of the slack variables needs to be minimized in the objective function, see (4-16). Using this formulation, a feasible solution can always be achieved.

$$\max \Delta g_{1,i}(t) - \Delta r_{1,i}(t) - \sum_{n=1}^{N-1}\left[\varepsilon_{n,i}^{S}(t) + \varepsilon_{n+1,i}^{T}(t)\right]$$

s.t.

$$\begin{aligned}
&\Delta r_{n,i}(t) - \Delta r_{n+1,i}(t) = S_{n,i}(t-1) - \varepsilon_{n,i}^{S}(t), && n \in \{1,...,N-1\}\\
&\Delta g_{n+1,i}(t) - \Delta g_{n,i}(t)\\
&\quad = T_{n+1,i}(t-1) - S_{n,i}(t-1) - \left[g_{n+1,i}(t-1) - g_{n,i}(t-1)\right] - \varepsilon_{n+1,i}^{T}(t), && n \in \{1,...,N-1\}\\
&\Delta g_{n,i}(t) - \Delta r_{n,i}(t) \le c_n - \sum_{p \in Z_{n,i}} \alpha^{Q}\left[Q_{n,p}^{\max}(t-1)/s_{n,p}\right] - g_{n,i}(t-1), && n \in \{1,...,N\}\\
&\Delta g_{n,i}(t) - \Delta r_{n,i}(t) + g_{n,i}(t-1) \ge 0, && n \in \{1,...,N\}\\
&\varepsilon_{n,i}^{S}(t), \varepsilon_{n+1,i}^{T}(t) \ge 0, && n \in \{1,...,N-1\}
\end{aligned} \tag{4-16}$$

2) Transition period

As we discussed above, if there is a change on the offset (i.e., $\Delta g_{n,i}(t) \neq 0$), a transition period (a few cycles) is needed in order to reach the new signal timing. If one wants to avoid the transition period and implement the change (i.e., $\Delta r_{n,i}(t)$) in the immediate next cycle , a new formulation (4-17) can be adopted by forcing $\Delta g_{n,i}(t) = 0$ in (4-16).

$$\max\left(-\Delta r_{1,i}(t) - \sum_{n=1}^{N-1}\left[\varepsilon_{n,i}^{S}(t) + \varepsilon_{n+1,i}^{T}(t)\right]\right)$$

s.t.

$$\begin{aligned}
&\Delta r_{n,i}(t) - \Delta r_{n+1,i}(t) = S_{n,i}(t-1) - \varepsilon_{n,i}^{S}(t), && n \in \{1,...,N-1\}\\
&\left[-\Delta r_{n+1,i}(t) + g_{n+1,i}(t-1)\right] - \left[-\Delta r_{n,i}(t) + g_{n,i}(t-1)\right]\\
&\quad = T_{n+1,i}(t-1) - \varepsilon_{n+1,i}^{T}(t), && n \in \{1,...,N-1\}\\
&-\Delta r_{n,i}(t) \le c_n - \sum_{p \in Z_{n,i}} \alpha^{Q}\left[Q_{n,p}^{\max}(t-1)/s_{n,p}\right] - g_{n,i}(t-1), && n \in \{1,...,N\}\\
&-\Delta r_{n,i}(t) + g_{n,i}(t-1) \ge 0, && n \in \{1,...,N\}\\
&\varepsilon_{n,i}^{S}(t), \varepsilon_{n+1,i}^{T}(t) \ge 0, && n \in \{1,...,N-1\}
\end{aligned} \tag{4-17}$$

Since both (4-16) and (4-17) are linear programs, they can be solved by any linear programming techniques. In practice, we would like to utilize the formulation of (4-12) to generate the solution first; if the solution makes the green duration of one intersection negative, (4-16) can then be adopted to obtain a feasible solution. In the following, a Forward-Backward Procedure will be introduced to solve (4-12) and (4-15). The procedure can be easily understood and makes the problem solvable even by hand.

4.4 Solution Method – A Forward-Backward Procedure (FBP)

Before introducing the solution method, we first investigate the uniqueness of solution to the proposed maximum flow model (4-12). Assume $\left[\Delta r_{1,i}^*(t),...,\Delta r_{N,i}^*(t),\Delta g_{1,i}^*(t),...,\Delta g_{N,i}^*(t)\right]$ is the optimal solution of (4-12), one can verify that $\left[\Delta r_{1,i}^*(t)+x,...,\Delta r_{N,i}^*(t)+x,\Delta g_{1,i}^*(t)+x,...,\Delta g_{N,i}^*(t)+x\right]$, $x \in R$, is also an optimal solution of Eq. (4-12) because it generates the same objective value and satisfies all the constraints as well. It is because that one can shift the offset of every intersection along the oversaturated route by the same amount (x in this case), and that will not change the internal kinematic traffic flow relations between intersections, such as queue formation and discharging. In other words, a unique solution for intersection offsets is missing for this problem. To make the optimal solution unique, a boundary condition $\Delta r_{1,i}(t)=0$ is added to the problem (4-12), indicating that we take the green starting point of the first intersection as the fixed reference point. Similar analysis also applies to the problem (4-15) for two intersecting oversaturated routes.

To solve the maximum flow based control model, a Forward-Backward Procedure (FBP) is proposed and described in the following. The FBP consists of two processes: a forward process, which is applied along the direction of traffic, and a backward process, which follows the opposite direction.

4.4.1 FBP for One Oversaturated Route

1) Forward Process (FP)

The forward process aims to eliminate both spillovers and residual queues by reducing red or increasing green of oversaturated phase without considering the constraints from other conflicting phases. The process is applied along the direction of flow and calculates the red and green changes for each oversaturated phase during time period t (i.e. $\Delta r_{n,i}^{F}(t)$ & $\Delta g_{n,i}^{F}(t)$). Note that the superscript "F" indicates the "Forward" process.

To eliminate spillover, we need to adjust the red time. As discussed, to make the solution unique, the red change for the first intersection is set to zero, i.e. $\Delta r_{1,i}^{F}(t) = 0$. For any other node, based on the relationship between spillover time and red changes (the first equation in (4-12)), one can derive the equation $\Delta r_{n,i}^{F}(t) = \Delta r_{n-1,i}^{F}(t) - S_{n-1,i}(t-1)$. So the red change for each node can be calculated by the following:

$$\begin{cases} \Delta r_{1,i}^{F}(t) = 0 \\ \Delta r_{n,i}^{F}(t) = \Delta r_{n-1,i}^{F}(t) - S_{n-1,i}(t-1), \quad n = 2,...,N \end{cases} \tag{4-18}$$

After determining the red time changes, we further adjust green time to eliminate residual queue. In order to find the maximum flow through the network, the flow out of the first node is set to its arc capacity, i.e., $\Delta g_{1,i}^{F}(t) = g_{1,i}^{a}(t)$. According to the second constraint in (4-12), we have the following equation, which generates the green changes for every node.

$$\begin{cases} \Delta g_{1,i}^{F}(t) = g_{1,i}^{a}(t) \\ \Delta g_{n,i}^{F}(t) = \Delta g_{n-1,i}^{F}(t) + T_{n,i}(t-1) - S_{n-1,i}(t-1) - \left[g_{n,i}(t-1) - g_{n-1,i}(t-1) \right], \ n = 2,...,N \end{cases}$$

(4-19)

Eq. (4-18) and Eq. (4-19) are very intuitive. Eq. (4-18) is to eliminate spillover by changing the duration of red lights. The amount of red time reduction at any intersection should accommodate not only the removal of the spillover to the upstream intersection ($S_{n-1,i}(t-1)$), but also the increase of the arrival flow due to the red reduction made at the upstream intersection ($\Delta r^F_{n-1,i}$). Eq. (4-19) is to extend green time by $T_{n,i}(t-1) - S_{n-1,i}(t-1) - \left[g_{n,i}(t-1) - g_{n-1,i}(t-1) \right]$ to discharge residual queues. Similarly, we need to account for the green change of the upstream intersection ($\Delta g^F_{n-1,i}$). The backward process will consider the situation if capacity constraints are violated.

2) Backward Process (BP):

The forward process follows the traffic direction and adds extra green time $\Delta g^F_{n,i}(t) - \Delta r^F_{n,i}(t)$ to discharge the residual queue and to remove spillover for each intersection. However, desired green increases for some intersections may not be achievable due to the other constraints, i.e., green time requirement for conflicting phases to discharge queues. To solve this problem, the backward process is designed to gate traffic when the required green time changes calculated in the forward pass are not feasible.

In this process, we start from the last intersection and follow the direction of the opposing flow to check whether the arc flow is less than or equal to arc capacity. The residual capacity $R_{n,i}(t)$ is calculated using Eq. (4-20) at each arc. Positive $R_{n,i}(t)$ means available green can accommodate the required green time increase $\left(\Delta g^F_{n,i}(t) - \Delta r^F_{n,i}(t) \right)$; negative $R_{n,i}(t)$ means available green is insufficient.

$$R_{n,i}(t) = g^a_{n,i}(t) - \left(\Delta g^F_{n,i}(t) - \Delta r^F_{n,i}(t) \right), n = N, ..., 1 \qquad \text{(4-20)}$$

After calculating the residual capacity for each arc, the backward green time adjustment $\Delta g_i^B(t)$ is equal to the minimum residual capacity for all arcs along the oversaturated route. Since $R_{1,i}(t) = 0$ ($\Delta g_{1,i}^F(t) = g_{1,i}^a(t)$, $\Delta r_{1,i}^F(t) = 0$), the minimum residual capacity is non-positive. If $\min_{n\in\{1,\ldots,N\}}\left[R_{n,i}(t)\right] = 0$, the requested green time increase $\left(\Delta g_{n,i}^F(t) - \Delta r_{n,i}^F(t)\right)$ from the forward process will be satisfied at all arcs and no further adjustment is needed in the backward process. However, if $\min_{n\in\{1,\ldots,N\}}\left[R_{n,i}(t)\right] < 0$, there is at least one arc where the capacity constraint is violated. The adjustment term $\Delta g_i^B(t)$ ($\Delta g_i^B(t) < 0$) is utilized to make sure the capacity constraints are satisfied at all arcs.

$$\Delta g_i^B(t) = \min_{n\in\{1,\ldots,N\}}\left[R_{n,i}(t)\right] \tag{4-21}$$

The final signal timing changes for every node are calculated by (4-22). The final green time change is equal to the summation of the calculated value in the forward process $\Delta g_{n,i}^F(t)$ and the adjustment term $\Delta g_i^B(t)$ in the backward process. Note that, $\Delta g_i^B(t)$ is the same for every intersection along the route.

$$\begin{cases} \Delta r_{n,i}(t) = \Delta r_{n,i}^F(t) \\ \Delta g_{n,i}(t) = \Delta g_{n,i}^F(t) + \Delta g_i^B(t) \end{cases}, \qquad n \in \{1,\ldots,N\} \tag{4-22}$$

4.4.2 FBP for a Network with Two Intersecting Oversaturated Routes

With two intersecting oversaturated routes, the available green at intersection I ($g^{a}_{I,\,i\&j}(t)$) needs to be split between phase i and j. We can split the available green time proportionally according to the requested green times from the forward process, i.e.,

$$\begin{cases} g^{a}_{I,i}(t) = g^{a}_{I,i\&j}(t) \times \dfrac{g^{F}_{I,i}(t)}{g^{F}_{I,i}(t) + g^{F}_{I,j}(t)} \\ g^{a}_{I,j}(t) = g^{a}_{I,i\&j}(t) \times \dfrac{g^{F}_{I,j}(t)}{g^{F}_{I,i}(t) + g^{F}_{I,j}(t)} \end{cases}$$

However, such split method may not be efficient because the binding constraints for available green times on one or both oversaturated routes may not come from the critical intersection. To overcome such deficiency, we can first compute the residual capacity for all intersections except the critical intersection I and, the backward adjustment term for both directions, $\Delta g^{B}_{i}(t)$ and $\Delta g^{B}_{j}(t)$, using equation (4-21). The "^" sign is used because here we did not consider intersection I. We can then calculate the requested green time increase for phase i and j of intersection I , denoted by $\Delta^{R}_{I,i}(t)$ and $\Delta^{R}_{I,j}(t)$.

$$\begin{cases} \Delta^{R}_{I,i}(t) = \Delta g^{F}_{I,i}(t) + \Delta g^{B}_{i}(t) - \Delta r^{F}_{I,i}(t) \\ \Delta^{R}_{I,j}(t) = \Delta g^{F}_{I,j}(t) + \Delta g^{B}_{j}(t) - \Delta r^{F}_{I,j}(t) \end{cases} \tag{4-23}$$

If $g^{a}_{I,\,i\&j}(t) \geq \Delta^{R}_{I,i}(t) + \Delta^{R}_{I,j}(t)$, then the available green constraint at intersection I is satisfied. The backward process adjustment terms $\Delta g^{B}_{i}(t)$ and $\Delta g^{B}_{j}(t)$ are equal to $\Delta g^{B}_{i}(t)$ and $\Delta g^{B}_{j}(t)$ respectively, see (4-24).

$$\begin{cases} \Delta g_i^B(t) = \Delta g_{\ i}^{\ B}(t) \\ \Delta g_j^B(t) = \Delta g_{\ j}^{\ B}(t) \end{cases} \tag{4-24}$$

Otherwise, the available green time at intersection I cannot satisfy the total requested green time increase for both directions i and j. The total available green time $g_{I,\,i\&j}^a(t)$ is split proportionally to two directions, $g_{I,\,i}^a(t)$ and $g_{I,\,j}^a(t)$ according to the requested green time increase, see (4-25). The total backward process adjustment terms for two directions are determined by (4-26).

$$\begin{cases} g_{I,i}^a(t) = g_{I,i\&j}^a(t) \times \dfrac{\Delta_{I,i}^R(t)}{\Delta_{I,i}^R(t) + \Delta_{I,j}^R(t)} \\ g_{I,j}^a(t) = g_{I,i\&j}^a(t) \times \dfrac{\Delta_{I,j}^R(t)}{\Delta_{I,i}^R(t) + \Delta_{I,j}^R(t)} \end{cases} \tag{4-25}$$

$$\begin{cases} \Delta g_i^B(t) = g_{I,i}^a(t) - \left(\Delta g_{I,i}^F(t) - \Delta r_{I,i}^F(t)\right) \\ \Delta g_j^B(t) = g_{I,j}^a(t) - \left(\Delta g_{I,j}^F(t) - \Delta r_{I,j}^F(t)\right) \end{cases} \tag{4-26}$$

The final red/green time changes can then be calculated using (4-22) for both directions.

.

4.4.3 Optimality Analysis of the FBP

We now prove that the solution generated by the FBP is the optimal solution for the MFP. The proof is divided into two portions: one is for a single oversaturated path and the other is for two intersecting oversaturated paths.

Theorem: The FBP provides optimal solutions to the maximum flow based signal timing control models (4-12) and (4-15).

Proof:

1) For one oversaturated route

Assume $\left[\Delta g_{1,i}^{*}(t),...,\Delta g_{N,i}^{*}(t),\Delta r_{1,i}^{*}(t),...,\Delta r_{N,i}^{*}(t)\right]$ is the signal control solution provided by FBP. We first check the feasibility of the solution. Based on (4-20), (4-21) and (4-22),

$$\begin{aligned}\Delta g_{n,i}^{*}(t)-\Delta r_{n,i}^{*}(t) &= \Delta g_{n,i}^{F}(t)+\Delta g_{i}^{B}(t)-\Delta r_{n,i}^{F}(t)\\ &= \Delta g_{n,i}^{F}(t)-\Delta r_{n,i}^{F}(t)+\min_{m\in\{1,...N\}}\left[R_{m,i}(t)\right]\\ &\le \Delta g_{n,i}^{F}(t)-\Delta r_{n,i}^{F}(t)+R_{n,i}(t)\\ &= g_{n,i}^{a}(t)\end{aligned}$$

Therefore, the third constraint in (4-12) is satisfied. Since (4-18) and (4-19) are essentially derived from the first two constraints of (4-12), it is easy to show that $\left[\Delta g_{1,i}^{*}(t),...,\Delta g_{N,i}^{*}(t),\Delta r_{1,i}^{*}(t),...,\Delta r_{N,i}^{*}(t)\right]$ also satisfies the first two constraints of (4-12). Therefore, the FBP generates a feasible solution to the model (4-12).

We now show the optimality of the FBP solution by contradiction. Assume that the signal control solution provided by the FBP is not optimal. Therefore, there must exist a control vector $\left[\Delta\tilde{g}$ which satisfies all constraints in (4-12) and $\Delta\tilde{g}$ $(t)+\Delta r_{1,i}^{*}(t)$.

Based on the definition of $\Delta g_{i}^{B}(t)$ in (4-21), there exists a node $k\in M$ such that

$$\Delta g_{i}^{B}(t)=R_{k,i}(t)=g_{k,i}^{a}(t)-\left(\Delta g_{k,i}^{F}(t)-\Delta r_{k,i}^{F}(t)\right)$$

which implies $\Delta g_{k,i}^{*}(t)-\Delta r_{k,i}^{*}(t)=\Delta g_{i}^{B}(t)+\left(\Delta g_{k,i}^{F}(t)-\Delta r_{k,i}^{F}(t)\right)=g_{k,i}^{a}(t)$.

Add Eq. (4-18) to Eq.(4-19):

$$\left(\Delta g_{n+1,i}(t)-\Delta r_{n+1,i}(t)\right)-\left(\Delta g_{n,i}(t)-\Delta r_{n,i}(t)\right)=T_{n+1,i}(t-1)-\left[g_{n+1,i}(t-1)-g_{n,i}(t-1)\right],$$
$$n\in\{1,...,N-1\}$$

Summing up the above equation from $n=1$ to $n=k-1$, we have:

$$\left(\Delta g_{k,i}(t)-\Delta r_{k,i}(t)\right)-\left(\Delta g_{1,i}(t)-\Delta r_{1,i}(t)\right)=g_{1,i}(t-1)-g_{k,i}(t-1)+\sum_{m=2}^{k}T_{m,i}(t-1):=H$$

The above equation implies that

$$\begin{aligned}\Delta\tilde{\ } \quad \tilde{\ } \quad \tilde{\ } \quad \tilde{\ }\\ &>H+\left(\Delta g^{*}_{1,i}(t)-\Delta r^{*}_{1,i}(t)\right)\\ &=\Delta g^{*}_{k,i}(t)-\Delta r^{*}_{k,i}(t)\\ &=g^{a}_{k,i}(t)\end{aligned}$$

This inequality violates the capacity constraint (available green) of arc *k* in model (4-12). Therefore, no solutions can be better than the one provided by FBP.

2) For two intersecting oversaturated routes

Assume $\left[\Delta g^{*}_{1,i}(t),...,\Delta g^{*}_{N,i}(t),\Delta r^{*}_{1,i}(t),...,\Delta r^{*}_{N,i}(t),\Delta g^{*}_{1',j}(t),...,\Delta g^{*}_{N',j}(t),\Delta r^{*}_{1',j}(t),...,\Delta r^{*}_{N',j}(t)\right]$ is the solution provided by FBP for two intersecting oversaturated paths. Similar approaches can be applied to show the solution satisfies the first three constraints of (4-15), since it is the same as (4-12). We just need to show the validity of the final constraint. According to (4-21),

$$\begin{aligned}&-\Delta r^{*}_{I,i}(t)-\Delta r^{*}_{I,j}(t)+\Delta g^{*}_{I,i}(t)+\ \Delta g^{*}_{I,j}(t)\\ &=-\Delta r^{F}_{I,i}(t)-\Delta r^{F}_{I,j}(t)+\Delta g^{F}_{I,i}(t)+\Delta g^{B}_{i}(t)+\ \Delta g^{F}_{I,j}(t)+\Delta g^{B}_{j}(t)\end{aligned}$$

If $g^{a}_{I,\,i\&j}(t)\geq\Delta^{R}_{I,i}(t)+\Delta^{R}_{I,j}(t)$, based on (4-23) and (4-24),

$$
\begin{aligned}
&-\Delta r^{*}_{I,i}(t)-\Delta r^{*}_{I,j}(t)+\Delta g^{*}_{I,i}(t)+\ \Delta g^{*}_{I,j}(t)\\
&=-\Delta r^{F}_{I,i}(t)+\Delta g^{F}_{I,i}(t)+\Delta g^{B}_{i}(t)-\Delta r^{F}_{I,j}(t)+\ \Delta g^{F}_{I,j}(t)+\Delta g^{B}_{j}(t)=\Delta^{R}_{I,i}(t)+\Delta^{R}_{I,j}(t)\le g^{a}_{I,\,i\&j}(t)
\end{aligned}
$$

If $g^{a}_{I,\,i\&j}(t)<\Delta^{R}_{I,i}(t)+\Delta^{R}_{I,j}(t)$, based on(4-25) and (4-26),

$$
\begin{aligned}
&-\Delta r^{*}_{I,i}(t)-\Delta r^{*}_{I,j}(t)+\Delta g^{*}_{I,i}(t)+\ \Delta g^{*}_{I,j}(t)\\
&=-\Delta r^{F}_{I,i}(t)-\Delta r^{F}_{I,j}(t)+\Delta g^{F}_{I,i}(t)+\Delta g^{B}_{i}(t)+\ \Delta g^{F}_{I,j}(t)+\Delta g^{B}_{j}(t)=g^{a}_{I,i}(t)+g^{a}_{I,j}(t)=g^{a}_{I,\,i\&j}(t)
\end{aligned}
$$

Therefore, the FBP for two intersecting oversaturated paths provides a feasible solution to the MFP (14).

We now show the optimality of FBP under two conditions: according to (4-23) and (4-24), If $g^{a}_{I,\,i\&j}(t)\ge\Delta^{R}_{I,i}(t)+\Delta^{R}_{I,j}(t)$, then the capacity constraint at the intersection *I* is not binding and the maximum flow through the network is determined by other intersections. Similar to the optimality proof for one path, the FBP for two intersecting path generates the optimal solution to the MFP (4-15).

If $g^{a}_{I,\,i\&j}(t)<\Delta^{R}_{I,i}(t)+\Delta^{R}_{I,j}(t)$, then the capacity constraint is binding at intersection *I*. The maximum total flow for two paths is constrained by the total available green time of intersection *I*. As long as the assigned green time is smaller than the requested green time, i.e., $g^{a}_{I,i}(t)\le\Delta^{R}_{I,i}(t)$ and $g^{a}_{I,j}(t)\le\Delta^{R}_{I,j}(t)$, the total flow will be maximum; otherwise, green time will be wasted. According to (4-25), one can derive

$$
\begin{cases}
g^{a}_{I,i}(t)=g^{a}_{I,i\&j}(t)\times\dfrac{\Delta^{R}_{I,i}(t)}{\Delta^{R}_{I,i}(t)+\Delta^{R}_{I,j}(t)}=\Delta^{R}_{I,i}(t)\times\dfrac{g^{a}_{I,i\&j}(t)}{\Delta^{R}_{I,i}(t)+\Delta^{R}_{I,j}(t)}<\Delta^{R}_{I,i}(t)\\
g^{a}_{I,j}(t)=g^{a}_{I,i\&j}(t)\times\dfrac{\Delta^{R}_{I,j}(t)}{\Delta^{R}_{I,i}(t)+\Delta^{R}_{I,j}(t)}=\Delta^{R}_{I,j}(t)\times\dfrac{g^{a}_{I,i\&j}(t)}{\Delta^{R}_{I,i}(t)+\Delta^{R}_{I,j}(t)}<\Delta^{R}_{I,j}(t)
\end{cases}
$$

Therefore, the FBP for two intersecting oversaturated paths generates one optimal solution to MFP (4-15).

4.5 Simulation Test

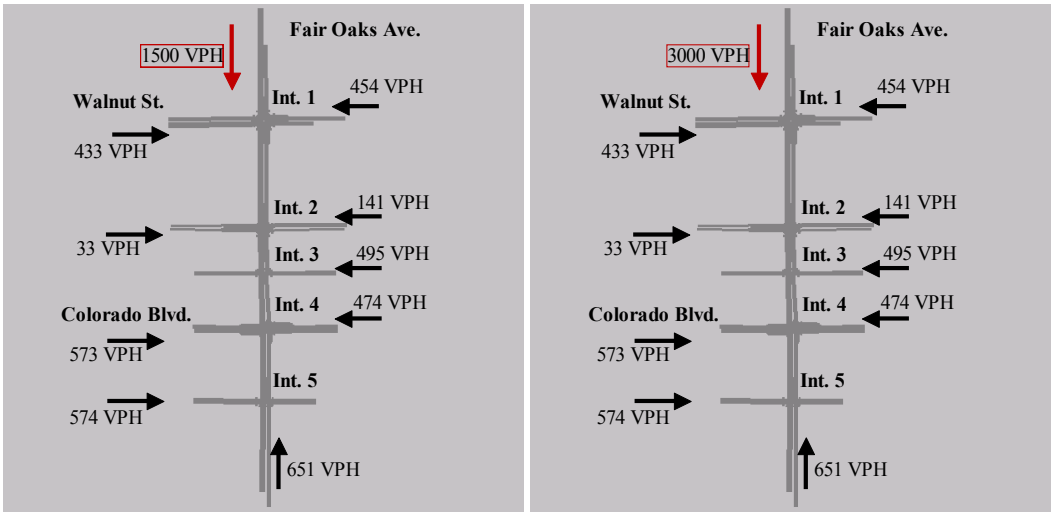
4.5.1 Simulation Design

A simulation study is conducted to test whether the proposed model can improve the performance of signalized arterials under oversaturated conditions. A network with 5 intersections along the Fair Oaks Ave in the City of Pasadena, CA is built in VISSIM (see Figure 4-11). The length of the corridor is 0.4 mile, the north-south direction is the coordinated direction and the speed limit is 30 MPH.

In order to test the performance of the proposed model, we create a southbound flow surge in the middle period of the simulation. The normal flow condition is shown in Figure 4-11 (a) with a southbound flow rate 1500 VPH and the increased flow condition is shown in Figure 4-11 (b) with a southbound flow rate 3000 VPH. As demonstrated by Table 4-1, during the first half an hour of the simulation (0~1800 Seconds), the flow condition is normal; the southbound flow rate is increased from 1500 VPH to 3000 VPH in the middle one hour (1800~5400 Seconds); finally, the network input flows get back to normal in the final half an hour (5400~7200 Seconds). The test scenario is designed to represent the traffic condition when there is a demand surge due to some unexpected reasons.

Synchro was first utilized to optimize the signal timings according to the normal and increased traffic flow conditions shown in Figure 4-11. With the normal flow, the optimized cycle length is 80 seconds; with the increased southbound flow, the optimized cycle length is 120 seconds. According to this, four types of control strategies are designed (see Table 4-2). Control strategy #1 and #2 represent the optimal signal timings according to normal flow and increased flow conditions respectively. Both of two

strategies are actuated-coordinated and they are implemented through the Ring-Barrier-Controller (RBC) in VISSIM. Control strategy #3 and #4 are essentially the same as Control Strategy #1 and #2, respectively, but follow the Forward-Backward Procedure proposed in this chapter with different cycle lengths (80 and 120 seconds) when oversaturated conditions are detected along the corridor. The control interval is set to 2 cycles. Since the signal timings need to be changed in real time for the FBP control to respond to the latest performance measures, i.e., TOSI and SOSI values, control strategy #3 and #4 are implemented through the COM interface in VISSIM.



(a) (b)

Figure 4-11 VISSIM simulation network (a) normal flow condition (b) increased flow condition

Table 4-1. Traffic flow conditions during the simulation period

Simulation time (Sec)	Traffic Flow Conditions
0~1800	Normal flow condition (a)
1800~5400	Increased flow condition (b)
5400~7200	Normal flow condition (a)

Table 4-2. Control strategy comparison

Control Strategy No.	Description	Cycle Length (Sec)
1	Actuated-coordinated control	80
2	Actuated-coordinated control	120
3	FBP	80
4	FBP	120

To compare the performance of different control strategies, the simulation was run 5 times with 5 different random seeds under each control strategy and the average performance of each control strategy over different random seeds is summarized below. One should note that, in real world, since it is difficult to obtain the real-time traffic demand information, it is almost impossible to change the signal timings to the corresponding optimal settings when demand changes. It is more likely and more realistic that when demand surge like Figure 4-11 (b) happens, the implemented control strategy in field will still be #1 (optimized according to the normal flow conditions). Therefore, it makes more sense to compare the performance of control strategy #1 and #3. However, for research purpose all the results under different control strategies are listed below and the performance of strategy #1 is considered as the base line.

4.5.2 Test Results

Table 4-3 and Figure 4-12 summarize the network performance during the whole simulation period under different strategies. For the average delay per vehicle, strategy #1 ends up with 81.37 seconds. With the optimal signal timings according to increased flow, strategy #2 reduces the number to 61.73 seconds, which is a 24.14% decrease. The FBP with a cycle length of 80 seconds reduces the average delay to 64.28 seconds (21% decrease), while the FBP with a cycle length of 120 seconds reduces the average delay to 56.95 seconds (30.02% decrease). For the average number of stops and average speed, similar trend can also be found where strategy #2, #3 and #4 substantially outperform strategy #1.

Table 4-3. Network performance comparison

	Strategy #1	Strategy #2		Strategy #3		Strategy #4	
		Value	(%)	Value	(%)	Value	(%)
Average Delay (Seconds/per veh.)	81.37	61.73	-24.14	64.28	-21.00	56.95	-30.02
Average # of stops (per veh.)	2.05	1.43	-30.34	1.60	-21.96	1.25	-39.12
Average Speed (MPH)	10.95	13.19	+20.42	12.92	+17.96	13.77	+25.76

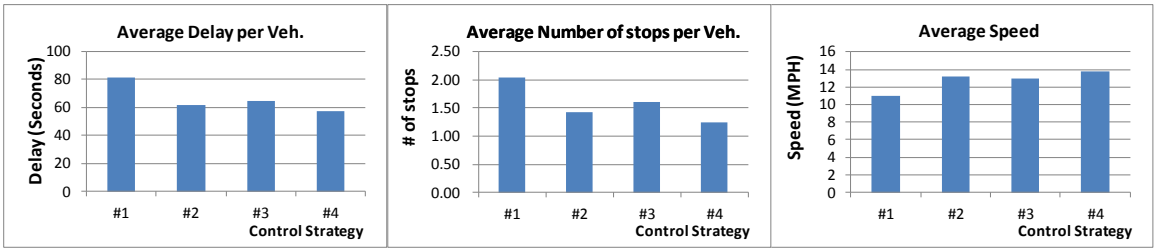


Figure 4-12 Comparison of network performance

Figure 4-13 compares the southbound throughputs over time under different control strategies. Each point represents the total throughputs during a 5-minute interval. One can see that, in the middle one hour period when the southbound input flow has a large increase, strategy #2, #3 and #4 have much better performance comparing with strategy #1 and within each interval more vehicles can be discharged through the southbound exit under strategy #2, #3 and #4. The total throughputs of different exits of the network (including all the side streets) in the 2-hour simulation period are summarized in Table 4-4. One can see that the throughput of the southbound exit for the two-hour period is 3021 vehicles under strategy #1. However, with strategy #2, #3 and #4 respectively, the total throughput is increased to 3716 (23.03% increase), 3762 (24.52% increase) and 3809 (26.10% increase). Overall, when comparing with the strategy #1, the total throughput of the whole network is increased by 9.0% under strategy #2, 9.6% under strategy #3 and 10.16% under strategy #4.

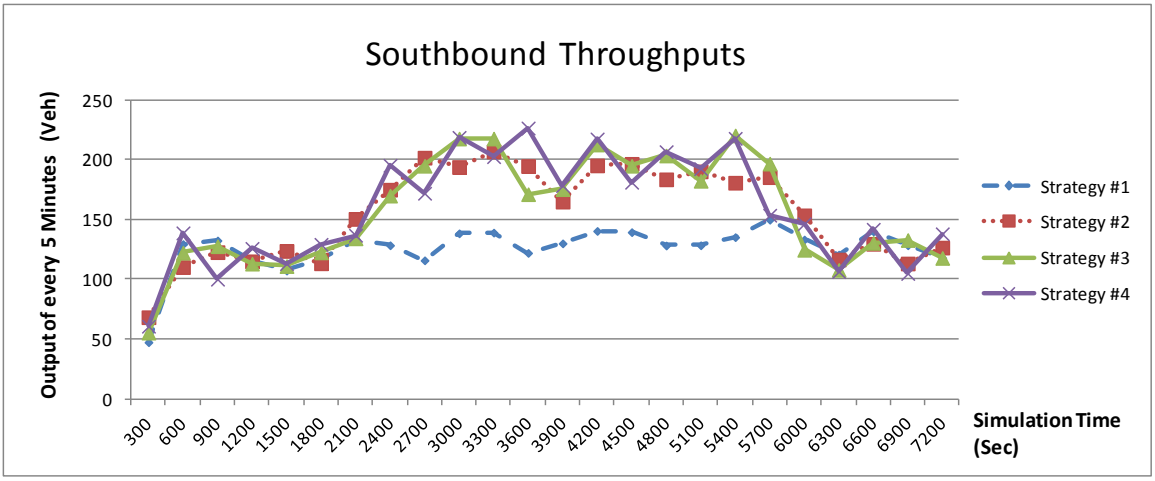


Figure 4-13 Southbound throughputs over time under different strategies

Table 4-4. Comparisons of two-hour throughputs

	Strategy #1	Strategy #2		Strategy #3		Strategy #4	
		Value	(%)	Value	(%)	Value	(%)
Southbound	3021	3716	+23.03	3762	+24.52	3809	+26.10
Northbound	1248	1244	- 0.32	1242	- 0.50	1246	- 0.18
Int. 1 side streets	1490	1546	+3.75	1539	+3.27	1544	+3.62
Int. 2 side streets	647	762	+17.73	772	+19.28	776	+19.93
Int. 3 side street	1120	1180	+5.36	1180	+5.32	1187	+5.93
Int. 4 side streets	1795	1800	+0.28	1815	+1.11	1806	+0.59
Int. 5 side street	1555	1609	+3.44	1613	+3.73	1614	+3.78
TOTAL	10880	11861	+9.0	11925	+9.6	11985	+10.16

Figure 4-14 compares the side streets' maximum queue length at two bottleneck intersections where the side street input demand is comparatively higher. One can see that, generally shorter cycle length generates shorter maximum queue length in each cycle. For example, strategy #1 has shorter maximum queue than strategy #2, and strategy #3 has shorter maximum queue than strategy #4. The maximum queue lengths on side streets in

each cycle under strategy #1 and #2 have less fluctuation than the ones under strategy #3 and #4. It is because that strategy #1 and #2 has the same cycle length, however, when the demand surge happens, strategy #3 and #4 will try to increase the discharging capacity of the oversaturated route by cutting some green time from side streets which may significantly increase the maximum queue length on side streets. But one can see from Figure 4-14, these increases are still within the acceptable ranges.

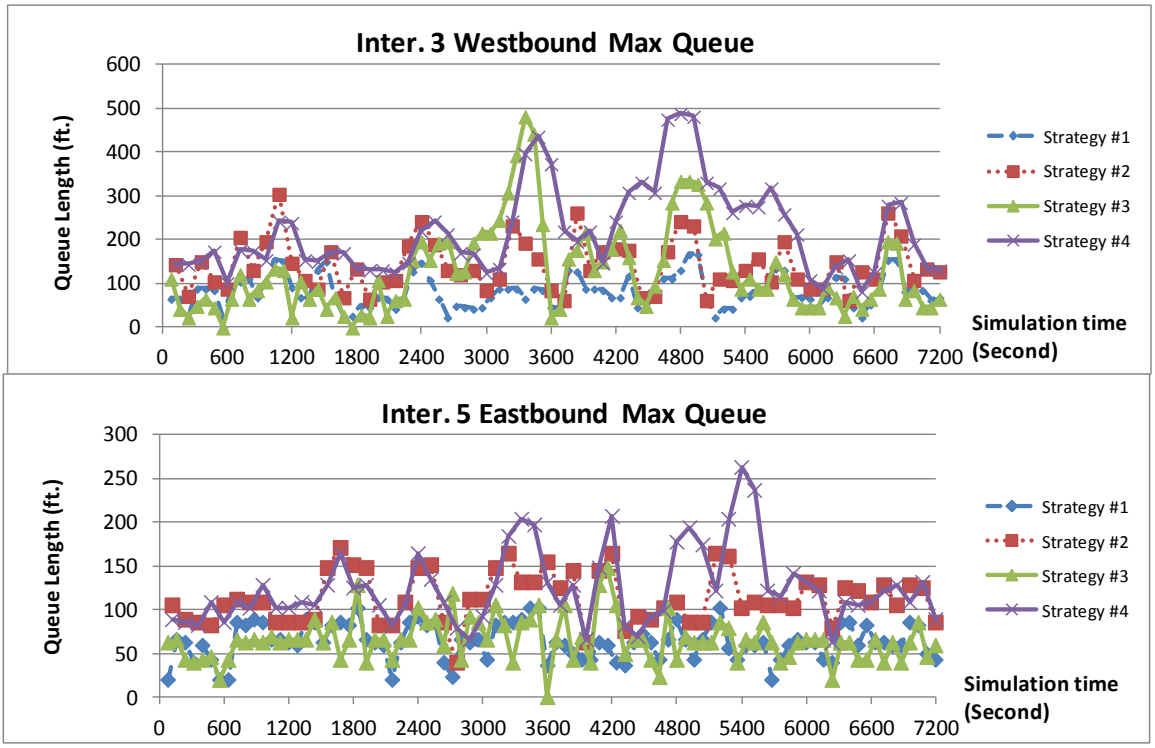


Figure 4-14 Comparison of side streets' maximum queue length in each cycle

To summarize, when oversaturation happens due to demand surge, strategy #2, #3 and #4 perform much better than the strategy #1 in terms of average delay per vehicle, average number of stops per vehicle, average speed and total throughputs. As discussed before, in real-world applications, real-time demand information is almost impossible to measure or estimate, especially when the demand is changing dramatically along time. However, the proposed FBP can adjust signal timings based on the measured Oversaturation Severity

Index and does not rely on the demand information. Therefore, it makes more sense to compare the performance of strategy #1 and #3. As one can see from the results above, the proposed FBP outperform Synchro in handling oversaturated traffic conditions, because the FBP systematically considers the discharging capacities between intersections and side street constraints along the oversaturated route. By applying the FBP, oversaturated conditions between intersections can be alleviated to the most extent, the traffic along the oversaturated route becomes much smoother and the total throughput can be significantly improved.

4.5.3 Sensitivity Analysis

In practice, the estimation of intersection TOSI and SOSI values inevitably has errors, which may impact the performance of the designed control model. In order to test the degree of influence, a sensitivity analysis is designed. In scenario 1, the estimated TOSI and SOSI values of all intersections are decreased by 10%; in scenario 2, all the values are increased by 10%. Everything else keeps the same as before. Table 4-5 compares the network performance of Strategy # 3 with variation to the estimation values. In general, the change of the estimated TOSI and SOSI values (i.e. decrease by 10% and increase by 10%) has minor impact to the network performance and the largest change is 3%. It indicates that the proposed control model is not sensitive to the estimation errors. One may also notice that the scenario with increased estimation values has better performance over the two, which implies that the original estimation may underestimate the TOSI and SOSI values.

Table 4-5. Network performance comparison between strategies with estimation errors

	Strategy #3	**Strategy #3** (Decrease estimation by 10%)		**Strategy #3** (Increase estimation by 10%)	
	Value	**Value**	**(%)**	**Value**	**(%)**
Average Delay (Seconds/per veh.)	64.28	64.76	0.75	62.30	-3.08
Average # of stops (per veh.)	1.60	1.64	2.50	1.58	-1.25
Average Speed (MPH)	12.92	12.86	-0.46	13.13	1.63

4.6 Summary

In this chapter, a simple but effective maximum flow based control model is developed to handle oversaturated traffic conditions at signalized arterials. The model does not rely on time-dependent traffic demand information; rather it is built upon the measurement of the oversaturation severity indices, which has been shown in Wu, et al. (2010). To solve the model, a Forward-Backward Procedure (FBP) is proposed in this chapter and it is mathematically proven that it generates the optimal solution to the model. The forward process aims to increase green time to mitigate oversaturation, therefore improve the throughput for the oversaturated approach; and the backward process aims to gate the traffic at some of the intersections to prevent residual queues and downstream queue spill-back when the available green time is insufficient. The FBP is very easy to understand, which makes the practical implementation of this model very promising. We have also tested the performance of the proposed model using microscopic traffic simulation in an arterial network in the City of Pasadena, CA. The results indicate that the proposed model significantly outperforms Synchro in handling oversaturated traffic conditions in terms of average delay per vehicle, average number of stops per vehicle, average speed and total throughputs.

5. An Integrated Control Model for Managing Network Congestion

In the previous chapter, a maximum flow based control model has been developed to manage the oversaturated signal arterials. It can be applied in many areas. In this chapter, we will develop an integrated control model to manage network congestion based on that.

5.1 Background

Integrated Corridor Management (ICM), which aims to leverage unused capacity within a corridor network, has drawn more and more attention in recent years because it is believed to be a promising tool to manage urban traffic congestion. The ICM-related research can date back several decades. In 1988, Van Aerde and Yagar (1988a and 1988b) first stated the importance of integrated control clearly and discussed the required characteristics to operate an integrated control system. Following that, researchers have established various integrated traffic control strategies, which can be roughly divided into two categories. The first category of strategies focuses on information provision and travelers' response, such as providing travel time information of different routes through Variable Message Sign (VMS). This group of study was first conducted by Papageorgiou (1990), who attempted to conceptually integrate ramp metering, real time information, route guidance and signal control for freeway corridor management. Their work was also extended by Hawas & Mahmassani (1995), Messmer & Papageorgiou (1995), Ben-Akiva et. al (1997), Pavlis & Papageorgiou (1999), Minciardi (2001), Wang & Papageorgiou (2002) and etc. These models aim to minimize the travel time difference from the origin node to the destination node through the available routes in the network and the decision variables of these models are splitting rates at each bifurcation node. However, most of these models require a known time-dependent traffic demand as inputs and that largely limits the practical application of this type of models. Further, even if the optimal

splitting rates can be calculated, since it is very difficult to estimate drivers' compliance rates, the actual performance would still be unknown. The second category of integrated traffic control strategies focuses on the traffic evolution and interaction between sub-systems. In 1995, Papageorgiou (1995) systematically developed an integrated control approach for traffic corridors including both freeways and signalized arterials based on the store-and-forward modeling philosophy. Later, Wu and Chang (1999) proposed a control model which integrates ramp metering, intersection signal timing, and off-ramp diversion under non-recurrent congestion. Liu et al. (2011) introduced a multi-objective optimization model to maximize the utilization of the available corridor capacity. However, the complexity of the models also limits the practical implementation, and even if they can be successfully solved, the ability of these models to deal with arterial traffic congestion still appears limited.

In this chapter, a new integrated control model is developed to reduce network congestion through diversion control. Comparing with the previous control models, the proposed one has the following merits:

(1) The impact of the diversion traffic to diverting route is specifically considered, especially for signalized arterials, so the potential congestion caused by diversion traffic can be reduced or eliminated by proper adjustment of signal timings.
(2) The proposed model does not have the requirement on time-dependent traffic demand information as model inputs. It is ready to be implemented at typical parallel traffic corridors where the standard detection system is available.
(3) The proposed model has very low computation burden and is suitable for on-line applications.

In the following, Section 5.2 defines the problem which will be solved in the research, followed by the detailed formulation of the proposed model in Section 5.3. The case study site and simulation results are presented in Section 5.4 and finally the summary of this chapter is given in Section 5.5.

5.2 Problem statement

This proposed model aims to solve the diversion control problem between two alternative routes in order to fully utilize the available capacities. A typical situation is that the two routes belongs to different control types, e.g. one route is freeway and the other is signalized arterial, as shown in Figure 5-1. The two origins O_1 and O_2 might be the same or different and so do the two destinations D_1 and D_2. In practice, most daily commuters would like to choose one of the routes based on their driving experience and preference. However, if the performance on one of the routes is significantly worse than the other, which might be caused by either recurrent (e.g. daily congestion during peak hours) or non-recurrent (e.g. car crash) event, to divert a portion of travelers to the alternative route with better performance would certainly benefit the whole system. Considering the diversion control between freeway and signalized arterial, when freeway congestion occurs, the control strategy is to divert the freeway traffic to the arterial system. How to inform travelers with real-time traffic information and how to predict the potential impacts of diverting traffic to the diverting route are the two most important questions which need to be answered in this chapter.

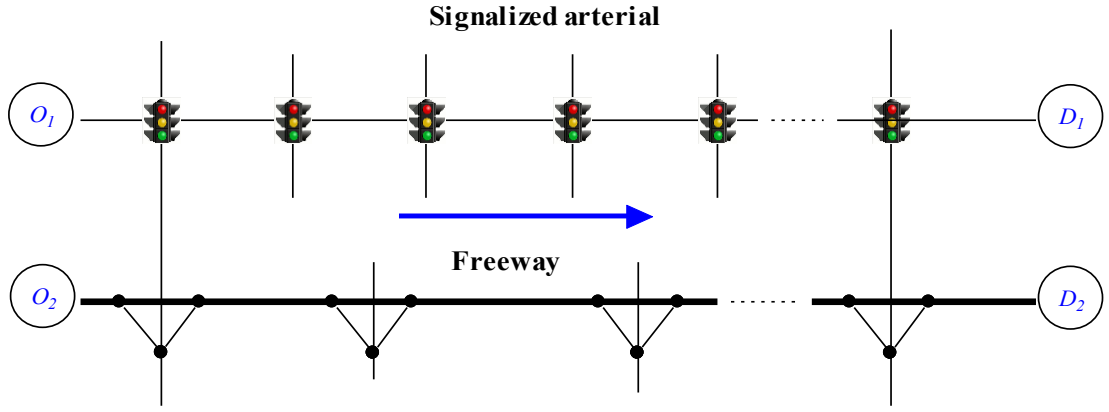


Figure 5-1 Problem statement

5.3 Model formulation

5.3.1 Performance estimation

In order to make correct control decisions, the performance of both routes needs to be monitored in real-time. At the end of each control period, control decisions for the next control period *t* will be made based on the traffic conditions in the immediately past control period *t-1*. The control interval usually includes 2~3 signal cycles. In this sub-section, the performance estimation method for both freeway and signalized arterial will be introduced.

1) Freeway performance estimation

Density and travel time are the two most important measures to reflect freeway performance. To estimate the real-time density and travel time on freeway, certain detection system (e.g. loop detectors, cameras, blue tooth technology and etc) is assumed to be available. Loop detector is one of the most commonly used techniques in the current traffic infrastructure. Detector stations are usually placed every 0.5 to 1 mile along freeways. The loop detector data, such as volume and occupancy, is transferred back to the control center in aggregated levels (e.g. every 30 seconds). In the proposed control model, a freeway corridor is divided into segments such that each segment contains one detector station. The performance of each segment is estimated based on the collected data from corresponding detector station.

Assume the freeway is divided into M segments, $1,2,...,M$. The density of each segment, denoted by $k_m(t-1)$ (Vehicles/Mile), can be calculated by (5-1), where $\theta_m(t-1)$ ($\theta_m(t-1) \in [0,1]$) is the average occupancy during the control period *t-1*, L_v is the average vehicle length (in feet) and L_d is the length of loop detector (in feet). Then, the average speed of each segment, denoted by $v_m(t-1)$ (Miles/Hour), can be generated by (5-2), where $V_m^{30}(t-1)$ is the average 30-second volume during control period *t*.

$$k_m(t-1) = \frac{5280 \times \theta_m(t-1)}{L_v + L_d}, m \in \{1, 2, ..., M\} \quad (5\text{-}1)$$

$$v_m(t-1) = \frac{120 \times V_m^{30}(t-1)}{k_m(t-1)}, m \in \{1, 2, ..., M\} \quad (5\text{-}2)$$

Thus, the travel time along the freeway corridor $T^f(t-1)$ can be calculated by (5-3), where l_m^f is the length for segment *m*. The superscript "*f*" represents freeway.

$$T^f(t-1) = \sum_{m=1,2,...M} \left[l_m^f / v_m(t-1) \right] \quad (5\text{-}3)$$

2) Arterial performance estimation

In order to estimate the arterial performance in real-time, the arterial traffic data collection system is also expected to be available, for instance, the SMART-Signal system (*Liu, et al., 2009*), which automatically archives the event-based high-resolution traffic data (i.e. signal changes and vehicle actuations). Based on the collected data set, real-time second-by-second queue length can be estimated with very high accuracy.

Assume the arterial has *N* signalized intersections and the queue length (in ft.) for the diverting traffic direction (i.e. phase *i*) at intersection *n* at any given second τ is $Q_{n,i}(\tau)$. It can be estimated using the method introduced by Liu, et al. (2009). For any cycle, the average delay $D_{n,i}$ (in Second/Vehicle) can be calculated by (5-4).

$$D_{n,i} = \frac{1}{A_{n,i} \times h} \sum_{\tau=1}^{c} \max\left[Q_{n,i}(\tau) - B_{n,i}(\tau), 0 \right] \quad (5\text{-}4)$$

Where *h* is the average space headway of vehicles in queue, *c* is the cycle length and $A_{n,i}$ is the total number of vehicle arrivals for phase *i* during the cycle. $B_{n,i}(\tau)$ is the distance from the front of the discharging wave to the stopbar at any given second τ. $B_{n,i}(\tau)$ is calculated by (5-5), where $\omega_{n,i}$ is the discharging wave speed (in ft./s) and $G_{n,i}$ is the green start time.

$$B_{n,i}(\tau) = \begin{cases} \omega_{n,i} \times (\tau - G_{n,i}) & \text{if } \tau > G_{n,i} \\ 0 & \text{else} \end{cases} \quad (5\text{-}5)$$

The average delay during control period *t-1*, denoted by $D_{n,i}(t-1)$, is the average of $D_{n,i}$ over all cycles within the period. Thus, the average travel time for signalized arterial during control period *t-1* is

$$T^a(t-1) = \sum_{n=1}^{N} \left(\frac{l_{n,i}^a}{v_{n,i}^a} + D_{n,i}(t-1) \right) \quad (5\text{-}6)$$

Where $l_{n,i}^a$ is the link length of approach *i* at intersection *n* along the signalized arterial and $v_{n,i}^a$ is the free flow speed. The superscript "*a*" represents signalized arterial.

Before making any adjustment, the residual capacity of each intersection needs to be calculated. The residual capacity of intersection n (denoted by $\eta_{n,i}(t-1)$) for the phase of diverting traffic direction (i.e. phase *i*) during the control period $t-1$ can be calculated by (5-7). $g_{n,i}(t-1)$ is the green time for phase *i* of intersection n during control period $t-1$, $s_{n,i}$ is the corresponding saturation flow rate, $z_{n,i}$ is number of lanes and $\gamma_{n,i}(t-1)$ is the

average cycle discharging volume for phase *i* of intersection n during control period $t-1$. The residual capacity measures how much more traffic can be discharged during one cycle at specific intersection based on the current traffic condition.

$$\eta_{n,i}(t-1) = g_{n,i}(t-1)s_{n,i}z_{n,i} - \gamma_{n,i}(t-1), n \in \{1,2,...,N\} \quad (5\text{-}7)$$

The residual capacity along the signalized arterial is the minimum residual capacity among all intersections,

$$\eta(t-1) = \min_{n\in\{1,2,...,N\}} \left[\eta_{n,i}(t-1)\right] \quad (5\text{-}8)$$

When the traffic along signalized arterial becomes congested, oversaturated traffic conditions may happen, which will cause detrimental effects to signal operation. An Oversaturation Severity Index (OSI) was proposed by *Wu et al. (2010)* to quantify the severity level of oversaturation by measuring its detrimental effects. Detrimental effect is characterized by either a residual queue at the end of a cycle or a spillover from downstream traffic, both of which create "unusable" green time. In the case of residual queue, the ''unusable" green time is the equivalent green time to discharge the residual queue in the following cycle, but for spillover, the ''unusable" green time is the time period during which an downstream link is blocked therefore the discharge rate is zero. OSI is further differentiated into *TOSI* (Temporal Oversaturation Severity Index, caused by the residual queue that creates the detrimental effect in temporal dimension) and *SOSI* (Spatial Oversaturation Severity Index, caused by the spillover that creates the detrimental effect in spatial dimension). In the following, we use $S_{n,i}(t-1)$ to represent the unusable green time caused by spillover at intersection *n* phase *i* during time period of $t-1$, and use $T_{n,i}(t-1)$ to represent the unusable green time caused by residual queue.

5.3.2 Diversion Control

Diversion control decisions are made based on the real-time estimated performance of both routes. When the performance on freeway is worse than that on arterial at control period $t-1$, we may want to divert some traffic from freeway to arterial, see Figure 5-2. This condition can be expressed by (5-9), where $T^{f\to a}(t-1)$ is the diversion cost from freeway to arterial, i.e. travel time on diverting links. In this case, a variable message sign (VMS) can be shown on the freeway side before the diverting point, indicating the travel times on both routes and advising drivers to use the arterial system.

$$T^{f}(t-1) > T^{f\to a}(t-1) + T^{a}(t-1) \tag{5-9}$$

Since there is usually no signal control on the freeway mainline, the exact number of diverting traffic is difficult to control. To overcome this problem, we assume drivers are rational and their perception errors follow a standard Gumbel distribution. Then, a Logit decision model is used to predict the diversion rate $\phi(t)$ at the next control period *t*, i.e. the percentage of vehicles that will be diverted from freeway to arterial because of the provision of traffic information of both routes. In the model, the diversion rate $\phi(t)$ is calculated based on travel time difference between arterial and freeway as shown in (5-10), where u is the travel time difference in minutes, α_2 is the coefficient that values travel time with respect to travel utility, and α_1 is the parameter that represents every other factor not related to time, such as drivers' inertia (i.e. unwillingness to divert). Both α_1 and α_2 can be estimated based on historical data and experience.

$$\begin{aligned} \phi(t) &= \frac{1}{1+e^{\alpha_1+\alpha_2 u(t-1)}} \\ u(t-1) &= T^{f}(t-1) - \left[T^{a}(t-1) + T^{f\to a}(t-1) \right] \end{aligned} \tag{5-10}$$

The diverted traffic volume from freeway to arterial in the next control period *t*, denoted by $\pi^{f\to a}(t)$, can be predicted by (5-11). The prediction is based on the assumption that the incoming traffic during control period t at the freeway is the same as that during control period $t-1$, denoted by $A^f(t-1)$. Because of the diverting traffic into the arterial system, the signal timings along arterial may need to be adjusted accordingly.

$$\pi^{f\to a}(t) = A^f(t-1)\phi(t) \tag{5-11}$$

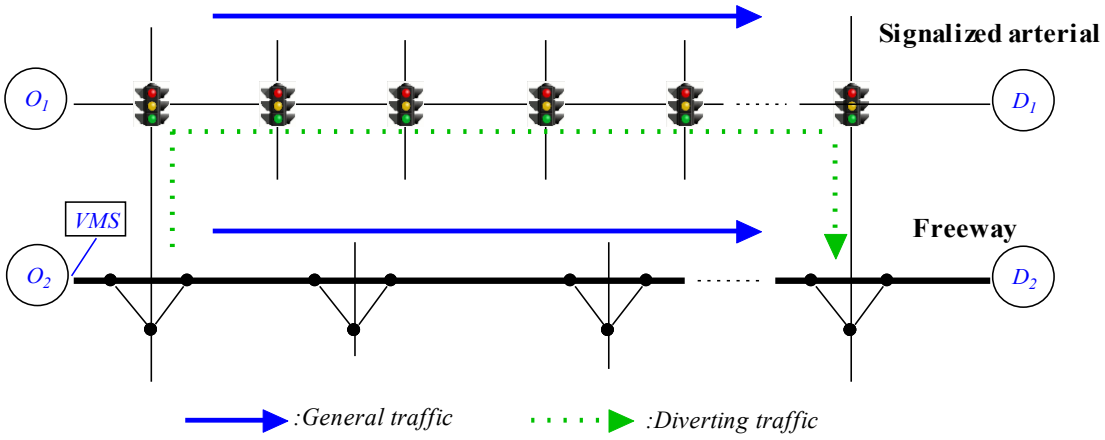


Figure 5-2 Diversion control from freeway to arterial

If $c\times\pi^{f\to a}(t)/\Delta t \le \eta(t-1)$, where *c* is the cycle length and Δt is the control interval, the diverting traffic can be handled by the current signal timings along the arterial; however, if $c\times\pi^{f\to a}(t)/\Delta t > \eta(t-1)$, the diverting traffic will cause residual queue at some intersection(s). The residual queue caused by the increase of traffic $\psi_{n,i}(t)$ at each intersection during the next control period *t* can be predicted by (5-12), where $\Delta\lambda_{n,i}(t)$ is the predicted increase of arrival traffic because of diversion control at intersection *n* during control period t. The boundary condition is

$\Delta\lambda_{1,i}(t) = c \times \left[\pi^{f \to a}(t) - \pi^{f \to a}(t-1) \right] / \Delta t$. The first equation basically says if the increase of arrival flow at specific intersection during control period t (i.e. $\Delta\lambda_{n,i}(t)$) is larger than the corresponding residual capacity (i.e., $\eta_{n,i}(t-1)$), there will be a residual queue $\Delta\lambda_{n,i}(t) - \eta_{n,i}(t-1)$; otherwise, there will be no residual queue at intersection *n*. The second equation updates the increase of arrival flow to the downstream intersection (i.e. $\Delta\lambda_{n+1,i}(t)$), which is equal to the minimum of the residual capacity (i.e., $\eta_{n,i}(t-1)$) and the increase of arrival flow (i.e. $\Delta\lambda_{n,i}(t)$) at the current intersection.

$$\begin{cases} \psi_{n,i}(t) = \max\left[0, \Delta\lambda_{n,i}(t) - \eta_{n,i}(t-1)\right] \\ \Delta\lambda_{n+1,i}(t) = \min\left[\eta_{n,i}(t-1), \Delta\lambda_{n,i}(t)\right] \end{cases}, n \in \{1, 2, ..., N\} \quad \text{(5-12)}$$

When residual queue happens at signalized intersections, it means the current discharging capacity cannot accommodate the increase of traffic. If the signal timings are not properly adjusted, more severe oversaturated conditions, such as spillovers, will appear. Therefore, the maximum flow based signal control model, which was introduced in Chapter 4, is utilized to mitigate or eliminate oversaturated traffic conditions between intersections. To accommodate the extra residual queue $\psi_{n+1,i}(t)$ caused by diversion control, Equation(4-12) is transformed into Equation(5-13). The same solution method still applies.

$$\max \Delta g_{1,i}(t) - \Delta r_{1,i}(t)$$

s.t.

$$\Delta r_{n,i}(t) - \Delta r_{n+1,i}(t) = S_{n,i}(t-1), \qquad n \in \{1,...,N-1\}$$

$$\Delta g_{n+1,i}(t) - \Delta g_{n,i}(t)$$
$$= T_{n+1,i}(t-1) + \psi_{n+1,i}(t) / \left(z_{n+1,i} s_{n+1,i}\right) - S_{n,i}(t-1) - \left[g_{n+1,i}(t-1) - g_{n,i}(t-1)\right], \qquad n \in \{1,...,N-1\}$$

$$\Delta g_{n,i}(t) - \Delta r_{n,i}(t) \le c_n - \sum_{p \in Z_{n,i}} \alpha^{Q} \left[Q_{n,p}^{\max}(t-1) / s_{n,p}\right] - g_{n,i}(t-1), \qquad n \in \{1,...,N\}$$

(5-13)

5.4 Case study and simulation

In order to test the proposed approach, a case study site was selected in Minneapolis, MN. As shown in Figure 5-3, there are two major routes, i.e. Trunk Highway 55 (a coordinated high speed signalized arterial) and Interstate freeway 394, connecting the west suburban living areas and the downtown Minneapolis. The total length of the corridor is about 3.5 miles and both routes (i.e. I-394 and TH 55) have a speed limit of 55 MPH. The coordination of the TH 55 favors the eastbound traffic during the AM peak hours because of the large traffic from home to work and it favors the westbound during the PM peak hours to handle the returning traffic. Based on the detector station locations in the field, the I-394 freeway is divided into 6 segments (see Figure 5-3) such that each segment contains one detector station. Figure 5-4 shows the flow-density diagram from the detector station at segment 4 based on the field collected data (the 30-second freeway data) between 6/15/2009 and 6/19/2009. One can easily find out that the critical density for segment 4 is about 105 Vehicles/Mile.

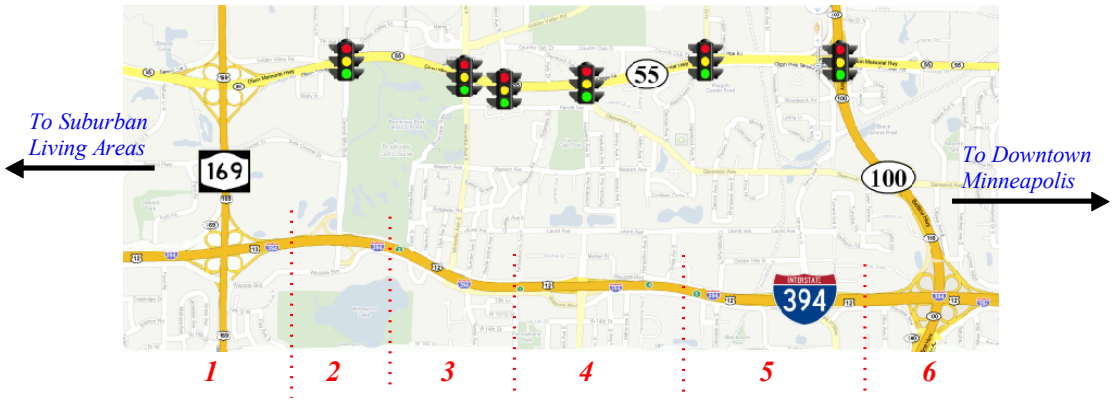


Figure 5-3 Case study site: the TH 55/I-394 corridor, Minneapolis, MN (Source: Google Maps)

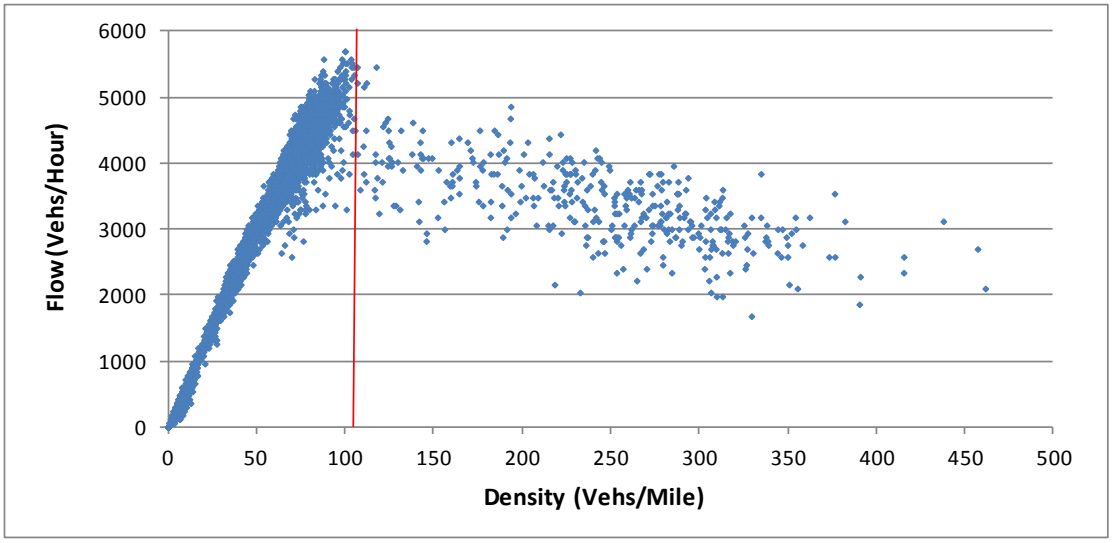


Figure 5-4 Flow-density diagram from three detectors at segment 4

A VISSIM model was built and calibrated using the field data collected during the morning peak hours (7:00 AM – 9:00 AM) between 6/15/2009 and 6/19/2009, see Figure 5-5. On I-394, the demands and turning percentages were estimated based on the 30-second data set archived by the freeway data collection system; on TH 55, the demands and turning percentages were generated by the SMART-Signal system. Morning peak

hours (7:00 AM – 9:00 AM) were simulated. Each testing scenario were run for 10 times using 10 different random seeds and the average results were used.

The diverting route is shown in Figure 5-5 by the green dotted line, which goes through TH 169 northbound, TH 55 eastbound and then TH 100 southbound. A simulation control program was written in C# and it controls the simulation in real-time through the COM interface of VISSIM. At each control period, the travel time on freeway (i.e. $T^f(t-1)$) and the travel time on the arterial (i.e. $T^a(t-1)$) is estimated based on the collected data from simulation; the diverting cost $T^{f\to a}(t-1)$ is the summation of travel times on TH 169 northbound and TH 100 southbound, which can also be estimated through the same approach as freeway.

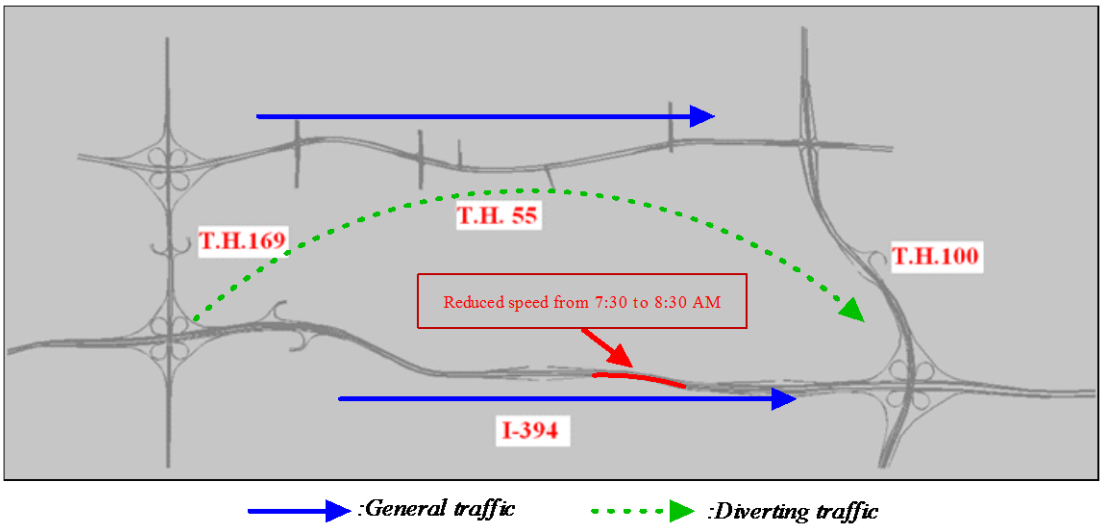


Figure 5-5 VISSIM network of the TH 55/I-394 corridor

Figure 5-6 shows the demand profiles of the major directions (i.e. I-394 EB, I-394 WB, TH 55 EB and TH 55 WB) for the whole simulation period. The cycle length of the signalized arterial is 180 seconds and the control interval Δt is 360 seconds. To simulate some unexpected incident (i.e. car crash) happening on freeway, a reduced speed area (10

MPH) with a length of 800 ft is created on the eastbound of I-394 from 7:30 AM to 8:30 AM (see Figure 5-5). Vehicles passing that area during that time window have to reduce their speed and as a result severe congestion will happen on the eastbound of I-394.

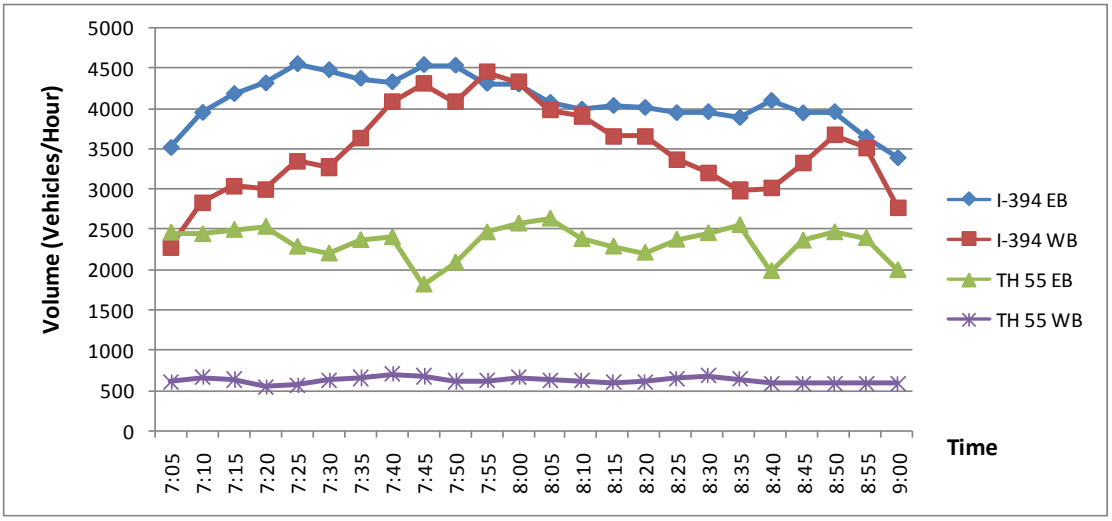


Figure 5-6 Demand profiles for the simulation period

In the following, two scenarios will be tested: one is the base scenario with the original control strategy (i.e. independent control) and the other is the scenario with the proposed integrated control strategy. Figure 5-7 shows the travel time profiles of the general route and diverting route. Under the base scenario with the original control strategy, the travel time of the general route (see the blue dashed line with diamond markers) increase dramatically after 7:30 AM when the congestion on freeway happens. The travel time of the diverting route (see the red dashed line with square markers) is relatively consistent during the whole period. Under the integrated control strategy, although the travel time of the general route (see the green line with triangle markers) still increases largely after 7:30 AM, the increasing trend is much slower because of diversion control; on the other hand, the diverting traffic makes the travel time of the diverting route (see the purple solid line with star markers) much higher comparing with that of base scenario.

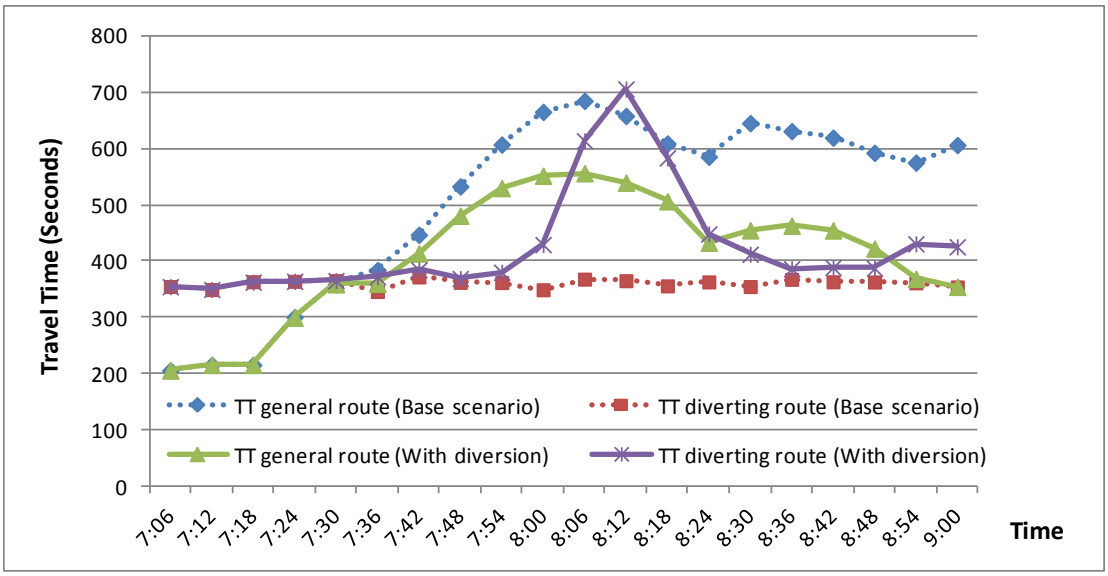


Figure 5-7 Travel times of general route and diverting route under different scenarios

Figure 5-8 presents the relationship between the travel times on different routes and the diversion rate. One can see that, at the beginning of the simulation, the travel time on the general route is lower than that on the diverting route, so no diversion control is needed; after the severe congestion happens on freeway, the travel time on the general route becomes higher than that on the diverting route, a portion of the traffic decides to use the diverting route, which will inevitably increase the travel time on the diverting route. The diversion control continues until the traffic situation reaches the point when the travel time on the diverting route becomes higher than that of the general route. And then, this process repeats. It is not difficult to find that the two lines (travel time on the general route and the diverting route) will weave with each other to best utilize the corridor capacities.

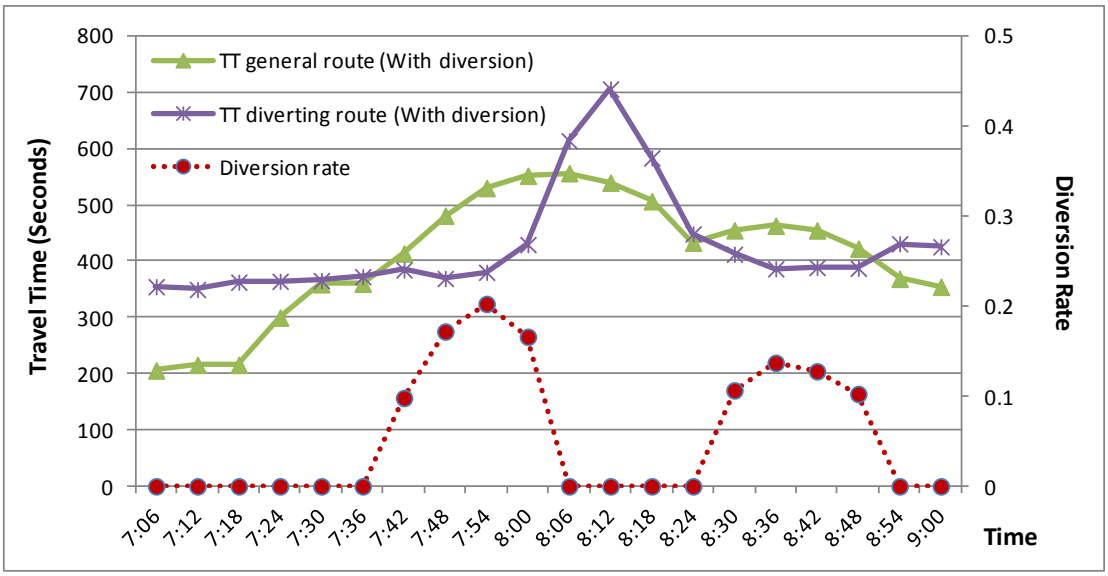


Figure 5-8 Travel time and diversion rate

Table 5-1 summarizes the network performance during the whole simulation period. With original control strategy, the average delay is 55.69 Seconds/Veh, while with the proposed integrated control strategy, the average delay is reduced to 41.14 Seconds/Veh, which is a 26.13% reduction. For average number of stops of the whole network, the proposed control model reduces it from 2.21 to 1.28, a 42.13% reduction. The average speed is increased from 42.12 MPH to 45.86 MPH.

Table 5-1. Network performance comparison

	Base Scenario	With diversion	Change
Average Delay (Seconds per veh.)	55.69	41.14	-26.13%
Average # of stops (per veh.)	2.21	1.28	-42.13%
Average Speed (MPH)	42.12	45.86	+8.89%

In order to test the performance of the proposed control strategy to handle different demand levels, we increase and decrease the mainline demand (i.e. the demand shown in Figure 5-6) by 5% and then run the simulation again. Table 5-2 presents the network performance under demand variations. When the mainline demand is increased by 5%, the whole network becomes more congested, which can be reflected by the increase of average delay and average number of stops and the decrease of average speed. However, with the proposed diversion control strategy, average delay and average number of stops can be reduced by 16.31% and 38.2% respectively and average speed can be increased by 7%. On the other hand, when the mainline demand is decreased by 5%, the proposed diversion control strategy can still significantly improve the network performance, i.e. reduce average delay by 29.67%, reduce average number of stops by 47.97% and increase average speed by 7.82%. Based on the results discussed above, one can see that the proposed integrated control model can effectively reduce network congestion and smooth traffic movement at different demand levels.

Table 5-2. Network performance comparison with demand variations

	Increase demand by 5%			**Decrease demand by 5%**		
	Base Scenario	**With diversion**	**Change**	**Base Scenario**	**With diversion**	**Change**
Average Delay (Seconds per veh.)	76.79	64.27	-16.31%	40.27	28.32	-29.67%
Average # of stops (per veh.)	3.45	2.13	-38.20%	1.41	0.73	-47.97%
Average Speed (MPH)	37.63	40.26	+7.00%	45.98	49.58	+7.82%

5.5 Summary

In this chapter, we propose an integrated control model to manage network congestion. Through diversion control, the model tries to fully utilize the available capacity along parallel routes. The impact of the diversion traffic is specifically considered, especially for signalized arterial, so the caused congestion can be reduced or eliminated by proper adjustment of signal timings. This model does not rely on time-dependent traffic demand as model inputs and it is ready to be implemented at typical parallel traffic corridors where the standard detection system is available. With the extremely low computation burden, the model is very suitable for on-line applications. We test the performance of the proposed model using microscopic traffic simulation in the I-394 and TH 55 corridor in Minneapolis, MN. The results indicate that the proposed model can significantly reduce the network congestion and make traffic much smoother, which can be reflected by the improvement on network performance measures, such as average delay per vehicle, average number of stops per vehicle and average speed.

6. Conclusion and Future Research

6.1 Conclusion

To tackle the congestion mitigation problem, this dissertation takes an important step to improve traffic operation and reduce congestion for signalized arterials. Based on the high-resolution traffic signal data collected by the SMART-Signal system and the derived performance measures, this research first proposes an innovative approach to optimize arterial offsets and then develops an effective method to manage oversaturated arterials. In particular, this research makes contribution in three major areas.

1) An offset optimization model for vehicle-actuated coordinated signals

The proposed model is developed based on the large amount of high-resolution field traffic data. It solves the two well-known problems with actuated signals: (1) the early return to green for coordinated phases and (2) the uncertainty of queue sizes formed at intersections. To account for the two problems, the concepts of conditional distribution of the green start times and traffic demand scenario are introduced. The queues formed by side-street and main-street traffic under different situations are explicitly considered. The objective of this model is to minimize total delay for one coordinated direction without sacrificing the performance of the other direction. The field test results indicate that the model successfully optimizes the offsets and significantly improves the performance of the corridor.

2) A maximum flow based approach to manage oversaturated arterials

A simple but effective maximum flow based control model is developed to manage oversaturated traffic conditions at signalized arterials. It is built upon the measurement of the oversaturation severity indices, which were introduced in Wu, et al. (2010). The model can be solved by the proposed Forward-Backward Procedure (FBP). The FBP is very easy to understand, which makes the practical implementation of this model very promising. The microscopic traffic simulation of an arterial network in the City of

Pasadena, CA shows that the proposed model significantly outperforms Synchro in handling oversaturated traffic conditions in terms of average delay per vehicle, average number of stops per vehicle, average speed and total throughputs.

3) An integrated control model for managing network congestion

The integrated control model tries to fully utilize the available capacity along parallel routes through diversion control. The impact of the diversion traffic is specifically considered, so the caused congestion can be reduced or eliminated by proper adjustment of signal timings. This model does not require time-dependent traffic demand as model inputs and it is ready to be implemented at typical parallel traffic corridors where the standard detection system is available. The model has extremely low computation burden, which makes it suitable for on-line applications. The microscopic traffic simulation in the I-394 and TH 55 corridor in Minneapolis, MN indicates that the proposed model can significantly reduce the network congestion and make traffic much smoother.

6.2 Future Research

This research makes an important step forward to improve the traffic signal operation using high-resolution data. The following is a list of topics that warrant further investigation:

1) Incorporate other signal control parameters

The proposed offset optimization model has shown its effectiveness to improve traffic progression at signalized arterials. It is meaningful to go further along this direction to consider other signal control parameters, such as green splits and cycle length. How the high-resolution data can help refine these parameters is a very important topic which needs to be considered in future research.

2) Field-test the oversaturation control strategies

In this research we make the first attempt to mitigate arterial oversaturation using measurable Oversaturation Severity Indices. Although we have demonstrated the effectiveness of the proposed model through simulation, it is necessary to validate the simulation results through field tests.

3) Expand to more complex road network

Our work in this dissertation is focused on a linear road network, i.e., an arterial or two intersecting arterials. However, in order to make the models more useful, the proposed models need to be extended for application on a more complex road network.

4) Consider more integrated control methods

For future research, more integrated control methods should be considered. For example, the freeway ramp metering control should be included in the model formulation to better control the traffic demand into the freeway system; on the other hand, variable speed limits could be another approach to smoothen the traffic on freeway. How to coordinate different control methods to improve network performance is worth further investigation.

References

Abu-Lebdeh, G. and Benekohal, R. F. (1997) Development of a traffic control and queue management procedure for oversaturated arterials. Transportation Research Record, 1603, 119-127.

Abbas, M., Bullock, D. and Head, L. (2001) Real-time offset transitioning algorithm for coordinating traffic signals. Transportation Research Record, 1748, 26–39.

Bavarez, E. and Newell, G.F. (1967) Traffic signal synchronization on a one-way street. Transportation Science, 1, 55-73.

Ben-akiva, M., Bierlaire, M., Bottom, J., Koutsopoulos, H. and Mishalani, R. (1997) Development of a route guidance generation system for real-time application. 8th IFAC Symposium on Transportation Systems, Chania, Creta, Greece

Chang, J., Bertoli, B. and Xin, W. (2010) New signal control optimization policy for oversaturated arterial systems. Proceedings of the 2010 Transportation Research Board Annual Meeting.

Chang, T. and Lin, J. (2000) Optimal signal timing for an oversaturated intersection. Transportation Research Part B, 34(6), 471-491.

Cohen, D., Head, L. and Shelby, S.G. (2007) Performance analysis of coordinated traffic signals during transition. Transportation Research Record, 2035, 19–31.

Day, C.M., Brennan, T.M., Hainen, A.M., Remias, S.M., Premachandra, H., Sturdevant, J.R., Richards, G., Wasson, J.S., and Bullock, D.M. (2011) Reliability, flexibility, and environmental impact of alternative objective functions for arterial offset optimization. Transportation Research Record, 2259, 8-22.

Day, C.M. and Bullock, D.M. (2011) Computational efficiency of alternative algorithms for arterial offset optimization. Transportation Research Record, 2259, 37-47.

Day, C.M., Haseman, R., Premachandra, H., Brennan, T.M., Wasson, J.S., Sturdevant, J.R. and Bullock, D.M. (2010) Evaluation of arterial signal coordination:

Methodologies for visualizing high-resolution event data and measuring travel time. Transportation Research Record, 2192, 37–49.

Federal Highway Administration (2007) "2007 National Traffic Signal Report Card", http://www.ite.org/reportcard/ accessed on 07/06/2011.

Federal Highway Administration (2012). Managing congestion with Integrated Corridor Management. http://www.its.dot.gov/icms/docs/cs_over_final.pdf, accessed on July 20, 2012

Gartner, N.H. (1983) OPAC: A demand-responsive strategy for traffic signal control. Transportation Research Record, 906, 75-81.

Gettman, D., Shelby, S.G., Head, L., Bullock, D.M., and Soyke, N. (2007) Data driven algorithms for real-time adaptive tuning of offsets in coordinated traffic signal systems. Transportation Research Record, 2035, 1-9.

Gazis, D.C. and R.B. Potts (1963) The oversaturated intersection. Proceedings of the Second International Symposium on the Theory of Traffic Flow, London, England, 1963.

Gazis, D.C. (1964) Optimal control of a system of oversaturated intersections. Operations Research, 12, 815-831.

Hawas, Y. and Mahmassani, H.S. (1995) A decentralized scheme for real-time route guidance in vehicular traffic networks. Intelligent Transport Systems World Congress, Yokohama, Japan

Head, K.L., Mirchandani, P.B. and Sheppard, D. (1992) Hierarchical framework for real-time traffic control. Transportation Research Record, 1360, 82-88.

Heydecker, B. (1987) Uncertainty and variability in traffic signal calculations. Transportation Research Part B, 21 (1), 79-85.

Hunt, P.B., Robertson, D.I., Bretherton, R.D.(1982) The SCOOT on-line traffic signal optimization technique. Traffic Engineering And Control, 23, 190-192.

Koonce, P., Rodegerdts, L., Lee, K., Quayle, S., Beaird, S., Braud, C., Bonneson, J., Tarnoff, P., and Urbanik, T. (2008) Traffic Signal Timing Manual, Federal Highway Administration Report: FHWA-HOP-08-024.

Lieberman, E., Chang, J., Prassas, E. (2000) Formulation of a real-time control policy for oversaturated arterials. Transportation Research Record, 1727, 77-88.

Little, J., Kelson, M. and Gartner, N. (1981) MAXBAND: A Versatile program for setting signals on arteries and triangular networks. Working paper 1185-81. Sloan School of Management, Massachusetts Institute of Technology.

Liu, H. and Ma, W. (2009). A virtual vehicle probe model for time-dependent travel time estimation on signalized arterials. Transportation Research Part C, 17(1), 11-26.

Liu, H., Wu, X., Ma, W. and Hu, H. (2009). Real-time queue length estimation for congested signalized intersections. Transportation Research Part C, 17(4), 412-427.

Liu, Y. and Chang, G. (2011) An arterial signal optimization model for intersections experiencing queue spillback and lane blockage. Transportation Research Part C, 19(1), 130-144.

Liu, Y. and Chang, G. and Yu, J. (2011) An Integrated Control Model for Freeway Corridor Under Nonrecurrent Congestion. IEEE Transactions on Vehicular Technology, 60, 1404-1418.

Lo, H.K., Chan, Y.C. and Chow, H.F (2001) A new dynamic traffic control system: Performance of adaptive control strategies for oversaturated traffic. Proceedings of the 2001 IEEE Intelligent Transportation Systems Conference, 404-409.

Luyanda, F., Gettman, D., Head, L., Shelby, S., Bullock, D. and Mirchandani, P. (2003) ACS-Lite algorithmic architecture: Applying adaptive control system technology to closed-loop traffic signal control systems. Transportation Research Record, 1856, 175-184.

MacCarley, C.A., Mattingly, S.P., McNally, M.G., Mezger, D. and Moore, J.E. (2002) Field operational test of integrated freeway ramp metering/arterial adaptive signal control: Lessons learned in Irvine, California. Transportation Research Record, 1811, 76-83.

Martin, P., (2007), Applications and benefits of adaptive traffic control systems in oversaturated conditions. Presentation at the 2007 TRB Workshop on Operating

Traffic Signal Systems in Oversaturated Conditions, Washington DC. Available online: http://www.signalsystems.org.vt.edu/documents/Jan2007AnnualMeeting/Presentations/ATCS_Oversaturation_TRB.pdf. Accessed on January 1, 2012.

Memon, G.Q. and Bullen, A.G.R. (1996) Multivariate optimization strategies for real-time traffic control signals. Transportation Research Record, 1554, 36-42.

Messer, C.J., Haenel, H.E. and Koeppe, E.A. (1974) Report on the User's manual for Progression Analysis and Signal System Evaluation Routine – PASSER II. Texas Transportation Institute Report 165-14.

Messmer, A. and Papageorgiou, M. (1995) Route diversion control in motorway networks via nonlinear optimization. IEEE Transactions on Control Systems Technology, 3 (1), 144-154

Michalopoulos, P.G. and Stephanopoulos, G. (1977a) Oversaturated signal system with queue length constraints – I: Single intersection. Transportation Research, 11(6), 413-421.

Michalopoulos, P.G. and Stephanopoulos, G. (1977b) Oversaturated signal system with queue length constraints – II: Systems of intersections, Transportation Research, 11(6), 423-428.

Minciardi, R. (2001) A decentralized optimal control scheme for route guidance in urban road networks. IEEE Intelligent Transportation Systems Conference Proceedings, Oakland, CA

Mulvey, J.M., Vanderbei, R.J. and Zenios, S.A. (1995) Robust optimization of large-scale systems. Operations Research, 43(2), 264-281.

Papageorgiou, M. (1990). Dynamic modeling, assignment, and route guidance in traffic networks. Transportation Research Part B, 24(6), 471-495.

Papageorgiou, M. (1995). An integrated control approach for traffic corridors. Transportation Research Part C, 3(1), 19-30.

Park, B., Messer, C. J. and Urbanik, T. (1999) Enhanced genetic algorithm for signal timing optimization of oversaturated intersections. Transportation Research Record, 1727, 61-67.

Park, B., Rouphail, N.M., and Sacks, J. (2001) Assessment of stochastic signal optimization method using microsimulation. Transportation Research Record, 1748, 40-45.

Pavlis, Y. and Papageorgiou, M. (1999) Simple decentralized feedback strategies for route guidance in traffic networks. Transportation Science, 33 (3), 264 – 278

Pignataro, L.J., McShane, W.R. and Crowley, K.W. (1978) Traffic control in oversaturated street networks. National Cooperative Highway Research Program Report 194. Transportation Research Board, Washington, DC

Rathi, A. K. (1988) A control scheme for high traffic density sectors. Transportation Research Part B, 22(2), 81-101.

Robertson, D.I. (1967) TRANSYT: A traffic network study tool. Report LR 253, Road Research Laboratory, Crowthorne, Berkshire, UK.

Sims, A. (1979) The Sydney coordinated adaptive traffic system. Presented at the ASCE Engineering Foundation Conference Research Priorities in Computer Control of Urban Traffic Systems, New York.

Skabardonis,A. (1996) Determination of timings in signal systems with traffic-actuated controllers. Transportation Research Record, 1554, 18-26

Shoup, G. and Bullock, D. (1999) Dynamic offset tuning procedure using travel time data. Transportation Research Record, 1683, 84-94.

Sussman, Joseph, et al. (2000) What Have We Learned About ITS? Report No. FHWA-OP-01-006. Federal Highway Administration, U.S. Department of Transportation. Washington, DC.

Trafficware. Synchro 5.0 User's Guide. Albany, California, 2001

Van Aerde, M. and Yagar, S. (1988a) Dynamic integrated freeway/traffic signal networks: Problems and proposed solutions. Transportation Research Part A 22(6), 435—443.

Van Aerde, M. and Yagar, S. (1988b) Dynamic integrated freeway/traffic signal networks: A routing-based modelling approach. Transportation Research Part A 22(6), 445—453.

Wallace, C.E., Courage, K.G., Reaves, D.P., Schoene, G.W. and Euler, G.W. (1984) TRANSYT-7F User's Manual. Transportation Research Center, University of Florida, Report No. UF-TRC-U32 FP-06/07.

Wang, Y. and Papageorgiou, M. (2002) A predictive feedback routing control strategy for freeway network traffic. Proceedings of the American Control Conference, Anchorage, AK

Wu, J. and Chang, G.L. (1999) An integrated optimal control and algorithm for commuting corridors. International Transactions in Operational Research 6(1), 39-55.

Wu, X., Liu, H. and Getttman, D. (2010) Identification of oversaturated intersections using high-resolution traffic signal data. Transportation Research Part C, 18(4), 626-638.

Yin, Y., Li, M.and Skabardonis, A., (2007) Offline offset refiner for coordinated actuated signal control systems, ASCE Journal of Transportation Engineering, 133(7), 423-432.

Zhang, L. and Yin, Y. (2008) Robust synchronization of actuated signals on arterials. Transportation Research Record, 2080, 111-119.